The Integration of Remotely Sensed Data with Stand-Scale Vegetation Models

Simon Ongeri

PhD

University of Edinburgh

1998

Acknowledgements

I would like to extend my sincerest thanks to my three supervisors, Profs. Paul Jarvis, John Monteith and Robert Gurney, for their boundless patience, advice and encouragement. Special thanks go out to Paul Jarvis for his expert guidance, for his extremely thorough editing of this thesis and for numerous discussions on all matters.

Thanks are also due to many other members of the IERM staff for their help, and for making my time in Edinburgh a very happy one. I would like to add particular thanks to Bart Kruijt for being such a good colleague, his help and his modelling wisdom, and to Shiela Wilson for things too numerous to mention.

Many thanks are also due to: David Pearson, Mike Barnsley, Peter North, and Jean-Cristophe Calvet for advice, collaborations and discussions, Agnes Bégué, John Moncrieff, Jon Lloyd, Philip Lewis and Peter Levy for the use of their HAPEX-Sahel data, the staff of the Edinburgh University Parallel Computing Centre for the use of their facilities, their advice and excellent training courses, and University of Cambridge Computer Science department for similar help.

Extra special thanks to Iain Beverland for being such a wonderful friend and colleague, his skills in cleaning blood stained shirts, and for his high speed driving skills that were responsible for getting me to hospital in record time.

Abstract

A study of vegetation modelling and remote sensing techniques revealed that it is not possible to measure many of the parameters required by mechanistic soil-vegetation atmosphere (SVAT) models using remote sensing. The physical distance between the sensor and target, the heterogeneity of the Earth's surface, the complex three-dimensional nature of vegetation canopies and the rapidly changing nature of the land-atmosphere system introduce significant errors into measurements. Because of these difficulties, it was concluded that remote sensing techniques are best used to study changes over time that result from growth processes in vegetation canopies. Active microwave (RADAR) remote sensing is a promising technique but sensors and platforms are still being developed, and, as a consequence, few data are currently available. Optical remotely sensed data are abundant but traditional methods for using the data (e.g. vegetation indices) are prone to difficulties. BRDF techniques offer several advantages over other optical techniques, but require the use of complex models of surface reflectance.

The vast majority of studies have attempted to derive vegetation model parameters from BRDF data, by inverting BRDF models using an inversion-parameterization strategy. In this study, a new validation-optimization strategy is proposed in which a SVAT and a *scene* model are linked. The output from the linked models is compared with remotely sensed data over time, and the model constrained to achieve the best fit. It is unlikely that remotely-sensed data will replace the need for ground-based measurements, instead the parameterization-optimization strategy uses remotely-sensed data to interpolate between ground-based measurements, enabling more accurate estimates of growth and yield.

The performance of four optimization techniques was compared, downhill simplex, direction set, simulated annealing and genetic algorithms. The downhill simplex and direction set methods, whilst effective on simpler problems, struggled to find global optima in the presence of local optima. The two stochastic methods, simulated annealing and genetic algorithms, are very powerful and were able to find global optima under extremely difficult conditions, however computationally they are both very intensive. Simulated annealing and genetic algorithms were successfully applied to both the inversion-parameterization and the validation-optimization strategies achieving excellent results in the optimization of BRDF models.

A linked SVAT-*scene* model has been developed that is able to combine remotely sensed data from a number of sources in a synergistic way. Object-oriented programming techniques were used to design the model, called FRAMEWORK, which is extremely flexible both in terms of its representation of the canopy (big-leaf or multi-layer), and in terms of the time step used. A suite of optimization techniques also forms part of the FRAMEWORK model structure, enabling the model to be optimized in a variety of ways. The model was validated using data for millet from the HAPEX-Sahel experiment in Niger. The model predictions of both canopy CO_2 fluxes and growth agreed well with measurements, with only small differences that could be attributed to uncertainties in canopy structural parameters, canopy heterogeneity and measurement and optimization noise. The results demonstrated that the combination of stochastic optimization techniques and the validation-optimization technique is a powerful tool for vegetation modelling.

The optimization of complex SVATs, with multiple canopy layers and short time-steps using stochastic optimization techniques is impractical using conventional computing techniques. The ability of high performance computing (HPC) techniques to overcome this problem was investigated by modifying FRAMEWORK to run on parallel computer architectures using the MPI language. The results indicate that parallel computing techniques can provide a successful way of improving the performance of complex SVAT models. The use of HPC with genetic algorithms also enables complex SVATS to be optimized by constraining them against remotely sensed data, in a manner that would not be possible using conventional computing methods

Contents

Chapter 1

Introduction

The prediction that human activities may lead to climate change is not a new one. The Swedish chemist, Svante Arrhenius, in 1896 foresaw that the release of carbon dioxide (CO_2) into the atmosphere as a result of burning fossil fuels would continue to increase ever more rapidly. This led him to predict that, as the amount CO_2 in the atmosphere doubled, the Earth would become several degrees warmer. What Arrhenius had envisaged, we now call the enhanced greenhouse effect. This is colloquially known as simply "the greenhouse effect", although this obscures the fact that the results of man-made emissions of gasses such as CO_2, merely enhance a natural effect, present in the Earth's atmosphere even before the industrial revolution.

The driving energy for the climate and weather comes from the sun. The Earth intercepts solar radiation, about one third of which is reflected, the remainder is absorbed by various components of the climate system, including, the atmosphere itself, oceans, land and biota. The outgoing radiation from the Earth consists of longwave radiation, the magnitude of which is determined by the temperature of the Earth-atmosphere system. Shortwave solar radiation passes through the clear atmosphere almost unimpeded, but longwave radiation emitted by the Earth's surface is absorbed and partially re-emitted by a number of trace gasses in the atmosphere, most notably water vapour and CO_2 (Figure 1.1). The effect occurs naturally in the atmosphere and is responsible for the Earth's surface being about 33 °C warmer than it would be if natural greenhouse gasses were not present. This is very fortunate for us all, since without the natural greenhouse effect life on Earth could not have evolved. In contrast the enhanced greenhouse effect resulting from human activities will probably bring about damaging changes in climate.

There is good historic evidence for the relationship between atmospheric CO_2 concentration and global mean temperature from ice core measurements. Figure 1.2 shows atmospheric CO_2 concentration and global mean temperature relative to the current values, obtained from isotopic analysis of ice cores from Vostock, Antarctica. A strong relationship exists between the isotopic concentration of Antarctic ice (deuterium or oxygen 18) and temperature. This relationship results from fractionation processes that take place during the atmospheric water cycle.

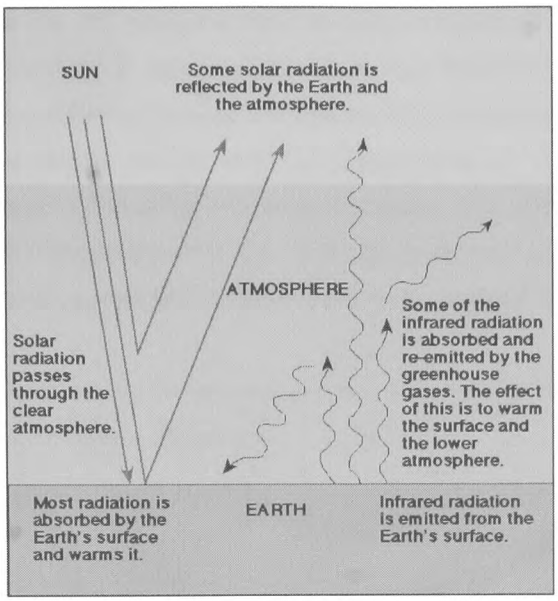


Figure 1.1 The greenhouse effect.

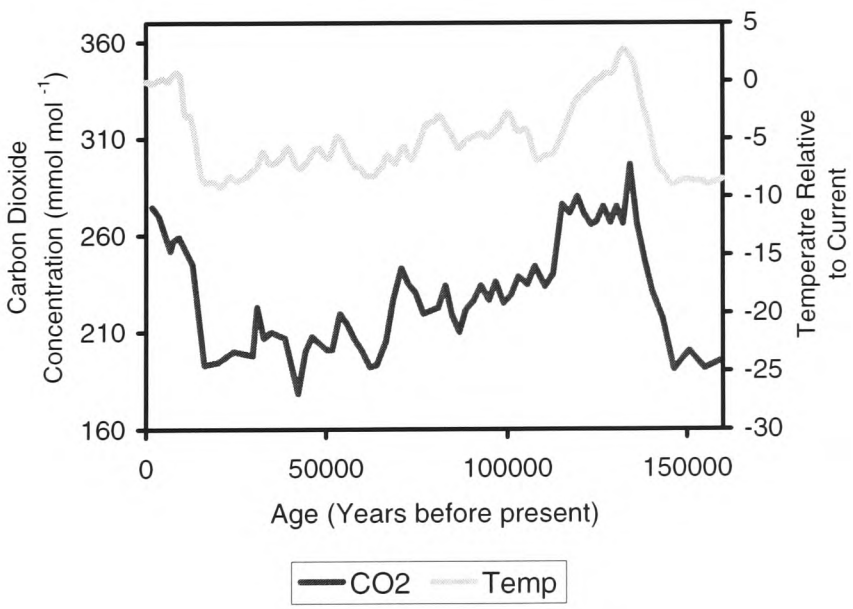


Figure 1.2 CO_2 concentration and estimated global temperature as determined from Vostock ice core data (Barnola *et al* 1987).

The ice core records indicate that, over the last 1000 years, the natural variability of atmospheric CO_2 concentration about the mean of about 280 μmol mol^{-1} was small (10 μmol mol^{-1} in 100 years). In

contrast, the rise over the last 150 years has been an order of magnitude greater, and the current concentration of 360 μmol mol^{-1} is unprecedented in the last 160,000 years. This information, combined with the strong correlation between atmospheric CO_2 increase and emissions from fossil fuel burning and land use change, and the observed isotopic trend of ^{13}C and ^{14}C, is compelling evidence of the anthropogenic nature of the increasing atmospheric CO_2 concentration. The observed rise in atmospheric CO_2 concentration is less than the estimated emission of carbon from anthropogenic sources, thus the oceans and terrestrial biota must, currently, be a sink for CO_2.

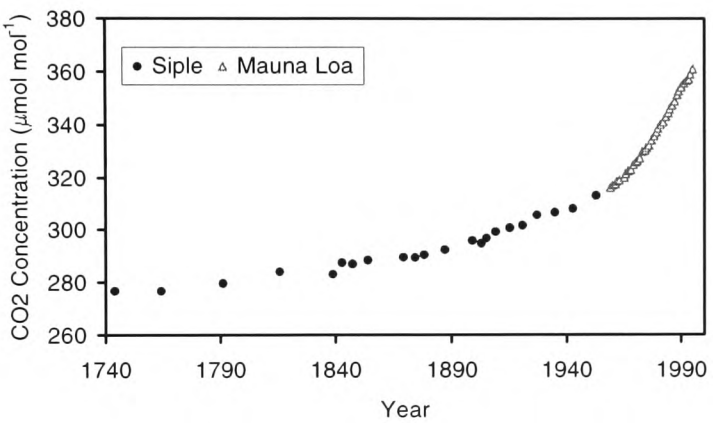


Figure 1.3 Atmospheric CO_2 increase over the past 250 years. Data from Siple ice core measurements (Antarctica), Neftel *et al*, 1985; and flask measurements at Mauna Loa (Hawaii), Keeling *et al*, 1989.

Tropospheric water vapour is the single most important greenhouse gas, but its atmospheric concentration is not significantly influenced by anthropogenic emissions. Of the greenhouse gasses that are directly influenced by human activities, CO_2 has the largest radiative effect. The pre-industrial CO_2 concentration was about 280 μmol mol^{-1} (derived from ice core data), corresponding to approximately 600 Pg of carbon. The current concentration is over 360 μmol mol^{-1}, and increasing at an ever more rapid pace, (Figure 1.3). The carbon in the atmosphere is not a static entity, but forms part of the global carbon cycle. Carbon in the form of CO_2, carbonates and organic material is continually cycled between various reservoirs in the atmosphere, oceans, land biota and marine biota. The exchanges between terrestrial biota and the atmosphere and the atmosphere and the surface water of the oceans account for by far the largest natural CO_2 fluxes. The net input into the atmosphere from fossil fuel burning and deforestation is much smaller in comparison, but large enough to affect the natural balance.

Figure 1.4 shows a schematic of the global carbon cycle, indicating the main reservoirs and the approximate magnitude of the annual fluxes. The largest global reservoirs of CO_2 are the oceans. There is a continuous exchange of CO_2 between the atmosphere and the oceans, driven by the difference between the partial pressure of CO_2 in the atmosphere and the equilibrium partial pressure in the surface waters. The exchanges between surface and deep-water layers shown in Figure 1.4 occur though oceanic circulation currents. The carbon cycle within the oceans is strongly influenced by the marine biota, which transport organic carbon from surface waters to deeper layers, in the form of a sinking current of detritus.

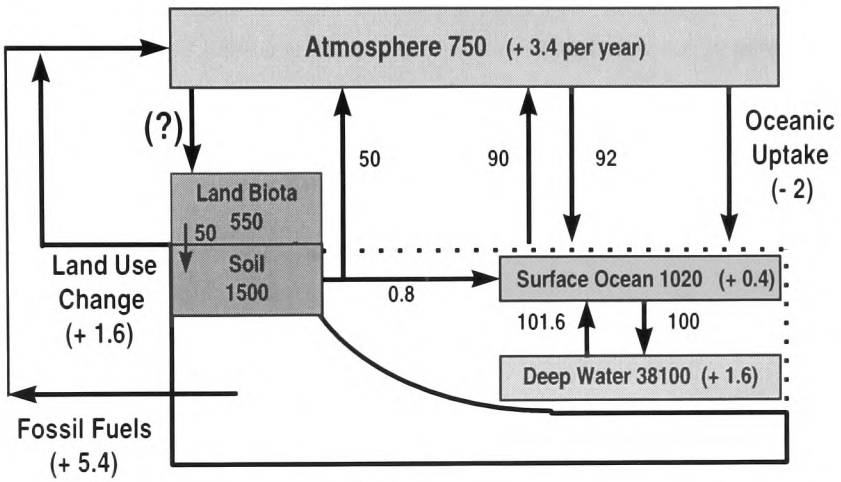


Figure 1.4 A simplified schematic diagram of the global carbon cycle. The numbers in boxes indicated the approximate magnitude of a carbon reservoir. The number is brackets indicate the approximate annual change, other numbers indicate annual fluxes. Units are Pg.

The most important processes involved in the exchange of CO_2 between the atmosphere and the land biota are photosynthesis, autotrophic respiration (CO_2 production by plants) and heterotrophic respiration (CO_2 production by microbes in soil). Net primary production (NPP) is the net annual CO_2 uptake by plants, and is equal to the gross uptake (GPP) minus autotrophic respiration. Before the interference of Man, NPP and decomposition by heterotrophic respiration were in approximate annual balance. However, Man may have exerted a considerable influence on this balance through the impact of land use change (e.g. deforestation) and climate change.

The main anthropogenic perturbations in the concentration of CO_2 in the atmosphere result from the burning of fossil fuels and changes in land use (for example deforestation). The global source of CO_2 into the atmosphere from fossil fuel burning and other minor sources such as cement production has shown an exponential increase since the second half of the nineteenth century.

The importance of terrestrial biota on the annual global carbon cycle is illustrated by Figure 1.5. The seasonal variations in CO_2 correspond to the action of photosynthesis by the terrestrial biota. The data show a more significant effect during the Northern Hemisphere summer because of the greater landmass in that hemisphere. In addition to their rôle as sinks for CO_2, vegetation has a further effect on climate: plants heat or cool the air around them through the reflection and absorption of solar radiation (sensible heat) and the evaporation of water vapour (latent heat). The effect plants have on the partitioning of radiant energy between sensible and latent heat flux is one to the most important factors in the development of the planetary boundary layer (PBL) which largely determines weather patterns near the ground. In turn, the pace and severity of changes in climate are likely to effect plant growth and function. Forests generally require cool conditions and may not be able to adjust to rapid climatic warming. However, a slower warming may allow forest boundaries to creep forward into more climatically favourable areas, leaving areas to be colonised by plants more suited to warmer conditions.

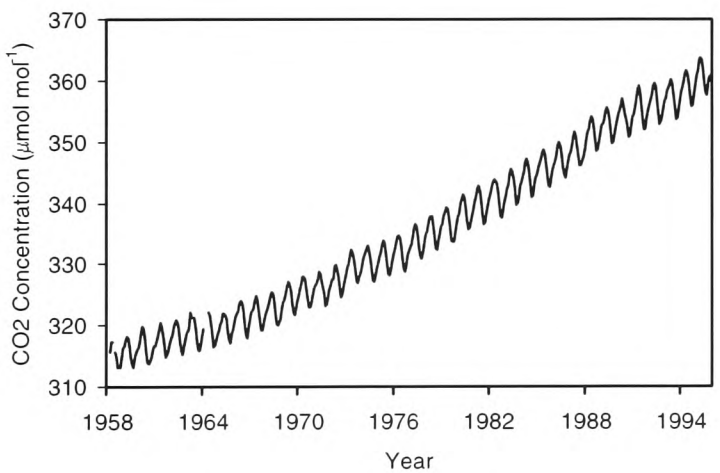


Figure 1.5 Atmospheric CO_2 concentration measured at Mauna Loa, Hawaii. The seasonal variations are primarily as a result of seasonal activity by terrestrial biota.

To be able to understand the likely progression of atmospheric CO_2 concentration and hence climate change, it is clearly of vital importance to understand the effects of vegetation on climate and the feedback effects changes in climate, such as temperature rise and increasing CO_2, have on vegetation. The rate of photosynthesis in plants is dependent upon the light intensity, temperature and availability of carbon dioxide, water and certain minerals. Elevated atmospheric CO_2 is therefore likely to result in increased rates of photosynthesis and plant growth in turn increasing the storage of carbon in plant tissue or soil organic matter and creating a negative feedback on atmospheric CO_2 increase. However, it is unclear over what time-scale this feedback may operate, or how it may be affected by other

factors. Increasing plant production may require an increasing availability of nutrients, which may ultimately limit any elevated CO_2 fertilisation effect. Increased CO_2 availability may reduce the transpiration of plants through a reduction in stomatal conductance, thus altering the partitioning of energy between sensible and latent heat. Both photosynthesis and respiration tend to increase with increasing temperature, but the increase in respiration with temperature is greater than that for photosynthesis. Consequently, a warming climate is likely to result, at least initially, in a release of CO_2 into the atmosphere. Increasing temperature may also result in increasing water use as plants attempt to control their temperature. Climate change may affect rainfall patterns leading to increase or reduced plant growth, possibly aggravating the effect of increasing temperature. The importance of the effect of climate change on vegetation, and any possible feedback effects that changes in vegetation might have on the future climate, are not yet fully understood.

Vegetation models are necessary for two reasons. Firstly, as a means of scaling up measurements, which usually cover small spatial and temporal scales, to the larger scales more relevant to the problems of global environmental change. Secondly, they allow predictions on how vegetation will behave under future climate scenarios. Because biological systems are so complex, mathematical models that describe these systems realistically are also complex. However, even the most complex models contain simplifying assumptions and one of the major goals of modelling is to determine the significance of the underlying assumptions they incorporate and evaluate the sensitivity of the models to these assumptions. One may, somewhat arbitrarily, divide models into "empirical" and "mechanistic" models. Empirical vegetation models make no attempt to describe the underlying processes involved, whereas mechanistic models attempt to explain processes at a more detailed level, but usually require far more detailed and complex parameterisation than empirical models.

To be able to apply models to the wide variety of different vegetation types and different regions of the world accurate parameterization is required. In addition, it is essential that models be validated against independent measurements. The determination of input parameters for mechanistic vegetation models (often called Soil Vegetation-Atmosphere Transfer schemes, or SVATs), such as MAESTRO (Wang and Jarvis 1991), is a complex, time consuming and expensive task. Typically, such models require meteorological conditions such as temperature, wind speed and humidity, vegetation physiology and structure, soil water and nutrient availability and incident radiation as input parameters. The dynamic nature of the Earth's environment means that these parameters are often rapidly changing. The collection of such datasets is obviously both costly and time consuming, and in some areas of the world may be completely impractical. The use of remote sensing techniques to parameterize and validate vegetation models would allow the application of models over wide areas of the Earth's surface.

It is desirable to be able to measure or monitor vegetation, remotely, without the need for direct human intervention, and without the cost of sending teams into the field to collect data. Remote sensing is defined as the acquisition of information about a target by a sensor that is not in physical contact with it. This term therefore could be applied to field measurements using radiometers, for example, however it is usually taken to mean measurement by sensors that are either flown on aircraft or satellites. Since much of this technology is continuously under development, many datasets have been acquired using truck mounted or airborne prototype instruments, in order to develop sensors for mounting on future satellite platforms. If techniques for integrating remotely sensed data with SVAT models could be found, they would enable our understanding of vegetation modelling to progress significantly, through the broadening of both the spatial and temporal extent of vegetation datasets. Whilst even a simple knowledge of physics is enough to tell us that parameters such as atmospheric humidity cannot be directly measured by remote sensing techniques, it may be possible to infer other parameters of interest such as standing biomass or absorbed radiation.

Remotely sensed data are also important for addressing the *scaling problem* – extrapolation of small-scale measurements or models to larger scales. The term *scale*, here, may refer to spatial or temporal scale. For example, in general circulation models (GCMs) of the climate, the Earth's surface is represented at a spatial scale of the order of 10,000 km. However, the energy budget used as an input into these models is usually measured by micrometeorological techniques over far smaller scales of the order of 100 to 1000 m. Extrapolation from small to larger scale is not straightforward for two reasons:

1) If the established response of a process at small scales is non-linear, then at a larger scale taking the linear average of the process will produce an incorrect result. Mathematically this is written:

$$y_i = f(x_i)$$

$$\bar{y} \neq f(\bar{x}) \qquad (1.1)$$

$$\bar{y} = \frac{\sum_{i=1}^{n} f(x_i)}{n}$$

2) At larger scales, new phenomena may emerge that are not considered at smaller scales. An excellent example of this in the context of SVATs is evaporation. At the leaf scale, evaporation can be described as a function of stomatal conductance and the atmospheric saturation vapour pressure deficit (VPD). However, at larger spatial scales, a community of leaves transpiring will

change the water content of the air sufficiently to reduce the VPD, thus providing a negative feedback on transpiration. At still larger scales, evaporation becomes more dependent on net radiation and entrainment into the atmospheric boundary layer (Jarvis and McNaughton 1986, McNaughton and Jarvis 1991).

Vegetation modellers are principally concerned with growth processes controlled by photosynthesis, and the exchange of fluxes between vegetation and the atmosphere. Photosynthesis is driven by the absorption of solar radiation and is a function of temperature, nitrogen and water availability. CO_2 diffuses into the reaction sites within the leaf, and oxygen diffuses out of the leaf, through stomata. Plants also transpire water vapour through stomata, in order to maintain a comfortable temperature by losing latent heat. The amount and arrangement of foliage, and the partitioning of radiation between diffuse and direct components, determine the light absorption of the vegetation canopy. Meteorological conditions such as wind speed and atmospheric humidity affect the rate of diffusion through leaf stomata, and thus influence the rate of photosynthesis, water loss through transpiration and evaporation, and the partitioning of energy into sensible and latent heat.

It is desirable to be able to measure these parameters directly from remotely sensed data. Other parameters, whilst not measurable directly by remote sensing, might be inferred from other quantities. Evapotranspiration is an important process as it is represents the water use of vegetation, and a measure of the partitioning of radiation between sensible and latent heat. Remote estimates of standing biomass would give an indication of the global storage of carbon in vegetation, or enable yield estimates for crops. Biomass is closely related to the vegetation leaf area index (LAI) which is a measure of the amount of foliage present. LAI is an essential parameter in many vegetation models, because it used to scale the amount of radiation absorbed by vegetation and the fluxes of water vapour and CO_2 between vegetation and the atmosphere. For natural systems, where there is no irrigation, water shortages can exert a significant influence over plant growth, therefore knowledge of soil water availability or plant vigour is once again very valuable.

1.1. AIMS AND OBJECTIVES

Remote sensed data are available at a range of spatial and temporal scales depending on the instrument and its flight plan. The use of these data would enable the development of area average descriptions of the soil-vegetation-atmosphere interface for use in large-scale models. The aims of this study was to investigate how SVAT models can be confronted with remotely sensed data. The

term "confronted" requires some explanation: it is used here to mean the use of remotely sensed data to parameterize, validate or constrain vegetation models. In the past, most studies have attempted to parameterize vegetation models by determining important parameters through empirical relationships with remotely sensed data, or by inverting models relating the remotely sensed signal to canopy properties. The success of these approaches has been very limited, therefore a number of objectives for this study were identified.

1) Determine which model parameters may be realistically derived from remotely sensed data, and to identify which of these parameters are the most important in terms of SVAT modelling.

2) Review previous studies where these parameters have been derived, and assess whether the techniques are applicable at the *stand-scale*.

3) To determine the most suitable remote sensing techniques and methods for confronting stand-scale models with data and determine a strategy for confronting models with remotely sensed data.

4) Devise new models where appropriate, and validate both the models and confrontation strategy using measured data.

Key to the fulfilment of these aims is an understanding of both SVAT modelling methods and remote sensing techniques. Therefore, a review of remote sensing techniques was undertaken, and the application of remote sensing to the derivation of parameters suitable for SVAT models was investigated. This study, described in Chapter 2, enabled existing strategies for using remotely sensed data to be assessed and new ones devised. The study revealed that it is not possible to measure many of the parameters required by stand-scale SVATs using remote sensing techniques. The physical distance between the sensor and target, the heterogeneity of the Earth's surface, the complex three-dimensional nature of vegetation canopies and the rapidly changing nature of the land-atmosphere system introduce significant errors into parameter derivations. Because of these difficulties, it was concluded that remote sensing techniques are best used to study changes over time that result from growth processes in vegetation canopies.

A review of the three major types of remote sensing, optical (visible and near-infrared wavelengths), active microwave (RADAR) and passive microwave techniques, identified optical and active microwave techniques as having potential for the derivation of stand-scale canopy information. Whilst good results may be obtained using active microwave remote sensing techniques (Chapter 2), instrument platforms are still being developed, and, as a consequence, datasets are not widely available. Optical remotely sensed data are abundant but traditional methods for using the data (e.g.

vegetation indices) are prone to signal saturation. New multi-angle techniques, such as measurement of the bidirectional reflectance distribution function (BRDF), offer several advantages over other optical techniques, but require the use of complex models of surface reflectance.

In recent years, a number of BRDF models have been developed that approximate the anisotropic reflectance of specific types of vegetation canopy (Chapter 3). Several of these models have been successfully inverted to retrieve canopy parameters from BRDF measurements. The complexity of BRDF models, and the number of parameters that must be retrieved in BRDF model inversion necessitate the use of relatively large measured datasets. This has usually resulted in inversion studies that utilise extensive canopy reflectance measurements that include data from a large number of different zenith and azimuth angles. The number of different directional measurements available from current aircraft and satellite remote sensing platforms is limited to a relatively small number of *looks* at a small range of zenith angles. However, new sensors are being developed that, although they still have a limited number of looks, are multispectral as well as multi-look instruments. The mechanical design of multi-spectral sensors is also much easier than multi-directional instruments, because the latter invariably need complex moving parts, such as mirrors, to point the sensor in the required direction. In general the successful inversion of complex BRDF models requires a larger amount of information (i.e. number of looks) than is available from directional measurements alone (Chapter 5). Although multi-spectral data are much easier to measure, the use of such data would require BRDF models with parameters that are independent of wavelength. This implies that models of leaf and soil optical properties (which are essential parameters of BRDF models) are required. Chapter 4 discusses the PROSPECT model (Jaquemoud and Baret 1990) of leaf optical properties and the SOILSPECT+ (Jaquemoud *et al.* 1992) soil reflectance model. In this study these two models were successfully combined with the SAIL BRDF model (Verhoef 1984) to form the PROSAIL+ model that contains only one wavelength dependent parameter, the single scattering albedo, ω. Chapter 4 also describes a technique for inverting PROSAIL+ using multi-spectral, multi-directional data, in which ω is tabulated for a range of different soil types, and instead of optimizing ω over a continuous range, the best fit value is extracted from a look-up table. This technique is used in later chapter to good effect.

The reflectance of a canopy at a given look angle is essentially a function of the leaf area index (LAI), leaf angle distribution (LAD), optical properties of the canopy elements, and the soil reflectance. Of these parameters LAI is the most important because it determines the amount of radiation absorbed by the canopy, and is closely correlated with biomass production. Attempts to relate the remotely sensed signal to other canopy properties are based on the assumption that the rate of change of canopy reflectance with the property is proportional to the rate of change of canopy reflectance with LAI (Chapter 2).

Although BRDF datasets provide more information about the surface than single-look data, the difficulties in deriving SVAT model parameters from these data are still large. Few BRDF models are mathematically invertible, consequently inversion techniques must employ a form of trial and error search. Moreover, BRDF models can only approximate the anisotropic reflectance of vegetation canopies. Usually simpler BRDF models are easier to invert, but give a less realistic representation of the canopy reflectance. The simplifying assumptions made in BRDF models are usually related to the type of vegetation canopy for which they were designed. A horizontally homogeneous canopy is a reasonable approximation for some crops such as wheat, however, a BRDF model with this assumption is unlikely to model the BRDF of forest canopies accurately. In addition, remote sensing measurements in the optical region are constrained by atmospheric effects such as clouds, which are omnipresent at many latitudes.

To circumvent many of the problems described above, a new strategy for confronting SVAT models with remotely sensed BRDF data is described in Chapter 5. The vast majority of studies have attempted to derive vegetation model parameters from BRDF data, by inverting BRDF models using an *inversion-parameterization* strategy. In this study, a new *validation-optimization* strategy is proposed in which a SVAT and a BRDF model are linked. The output from the linked models is compared with remotely sensed data over time, and the model constrained to achieve the best fit.

Constraining complex models requires the use of powerful optimization techniques. In Chapters 6 and 7 the performance of four optimization techniques (downhill simplex, direction set, simulated annealing and genetic algorithms) are compared using a particularly difficult function with many local optima. SVAT models are usually designed for a specific vegetation type (for example forests or crops) and the approximations encapsulated within the model reflect this fact. This lack of generality in models is further reflected in the way that individual sub-processes are not readily exchanged (for example, to replace an empirical model of photosynthesis with a mechanistic one) within the overall model structure (see for example MAESTRO, Wang and Jarvis 1988). In Chapter 8 the development of a new SVAT model, FRAMEWORK, is described, that is much more flexible it its structure. FRAMEWORK uses object oriented programming techniques to provide the intercommunications infrastructure between a number of modules. These modules represent various aspects of vegetation function such as photosynthesis and transpiration. The model design enables these modules to be changed or replaced easily enabling the model to be configured to almost any vegetation type, by combining different modules in different ways.

An optimization module forms an integral part of FRAMEWORK, enabling both the inversion-parameterization and validation-optimization approaches to be used to optimize the model. The flexible structure of FRAMEWORK and integrated optimization module allow the model to be confronted with remotely sensed data from a number of sources. In this study a bi-directional reflectance module is used, both to derive absorbed radiation and to predict canopy anisotropic reflectance and an optimization module. In Chapter 8, FRAMEWORK is validated in its different configurations, against datasets from the HAPEX-SAHEL project in Niger, West Africa.

Although stochastic optimization techniques are extremely powerful, the optimization of complex SVATs, with multiple canopy layers and short time-steps using such techniques, is impractical using conventional computing techniques. Chapter 9 describes the adaptation of FRAMEWORK to run on parallel computer architectures using the MPI language. In addition, Chapter 9 also describes the implementation of a parallel genetic algorithm optimization technique, which forms part of a parallel version of FRAMEWORK called P-FRAME. The performance gains achieved with P-FRAME are discussed in terms of the future of vegetation modelling.

Chapter 2

Review of Remote Sensing Techniques and Applications

Vegetation is a key component of terrestrial ecosystems, providing the basis of global food production through photosynthesis. Vegetation canopies also interact closely with the lower layer of the atmosphere, known as the planetary boundary layer, through the exchange of sensible and latent heat fluxes, momentum and trace gasses. Models of the interaction of vegetation with the environment called SVATS (Soil Vegetation Atmosphere Transfer Schemes) tend, by necessity, to be highly complex, because of the inherent complexity of natural systems. Models usually require many input parameters, which are difficult and costly to measure. The dynamic nature of the Earth's environment also means that these parameters are often rapidly changing. The use of remote sensing techniques to provide parameterizations for SVATS would theoretically allow rapid, automatic and repeated data collection over large, potentially inaccessible areas of the Earth, at a range of spatial scales. This would vastly increase the availability of data concerning terrestrial ecosystems. Unfortunately the difficulties encountered when attempting to use remote sensing techniques for vegetation monitoring are great. The relatively large distance between the sensor and target poses a problem with spatial resolution; temporal resolution is limited by the physics of satellite orbits or the availability and cost of aircraft flights. Moreover the Earth's surface is highly heterogeneous and the interaction of electromagnetic radiation with the land-atmosphere system is very complex.

A study of different remote sensing techniques and their potential for deriving canopy properties useful for vegetation modelling had been carried out and is the subject of this chapter. Remote sensing can be divided into two broad areas depending upon the range of wavelengths of the electromagnetic spectrum that the sensors use. These are *optical* remote sensing, concerned with the visible (VIS) and near infrared (NIR) part of the spectrum (approximately 400 nm to 700 nm for visible, and 700 nm to 1100 nm for NIR), and *microwave* remote sensing covering the 1mm to 100 mm range. In the past, most research and development has concentrated on optical platforms, probably because these wavelengths contain the part of the spectrum that is visible to humans and has consequently been extensively studied over hundreds of years. Unfortunately, clouds are opaque at VIS and NIR wavelengths and at many latitudes clouds obscure the surface for a large amount of the time making data acquisition (known as scenes) difficult.

2.1. VISIBLE AND INFRARED WAVELENGTHS

The absorption, scattering and reflection of electromagnetic radiation by a vegetation target modulate the characteristics of the emerging radiation. This modulation may be in the form of changes in the direction, brightness, polarisation and spectral properties of the radiation. It is this modulation which is used to infer properties about the vegetation. For example, the ratio of the reflected radiation in the NIR and VIS has been related to the photosynthetic capacity of vegetation (Tucker and Sellers 1986).

2.1.1. Atmospheric Correction

In the optical wavebands, the effect of the interaction between radiation reflected from the target and radiation scattered from atmospheric constituents between the instrument and the target cannot be ignored. Corrections must be made to the raw sensor signal to allow for this atmospheric effect. Atmospheric gasses and aerosols, together with clouds, absorb and scatter solar radiation, and can modulate the radiation reflected from the target through attenuation, altering its spectral distribution, or by introducing into the field of view sunlight scattered in the atmosphere: the larger the optical path between the sensor and target the larger the potential effect. Consequently, atmospheric corrections are vital for satellite platforms. The spectral bands used by satellite systems are usually located in atmospheric transmission windows, which are spectral regions where gaseous absorption is a minimum.

Figure 2.1 shows that in the visible part of the spectrum, atmospheric transmission is influenced mainly by molecular scattering and absorption by ozone. In the NIR there is a strong, narrow oxygen absorption band at 0.76 μm and a few water and carbon dioxide absorption bands.

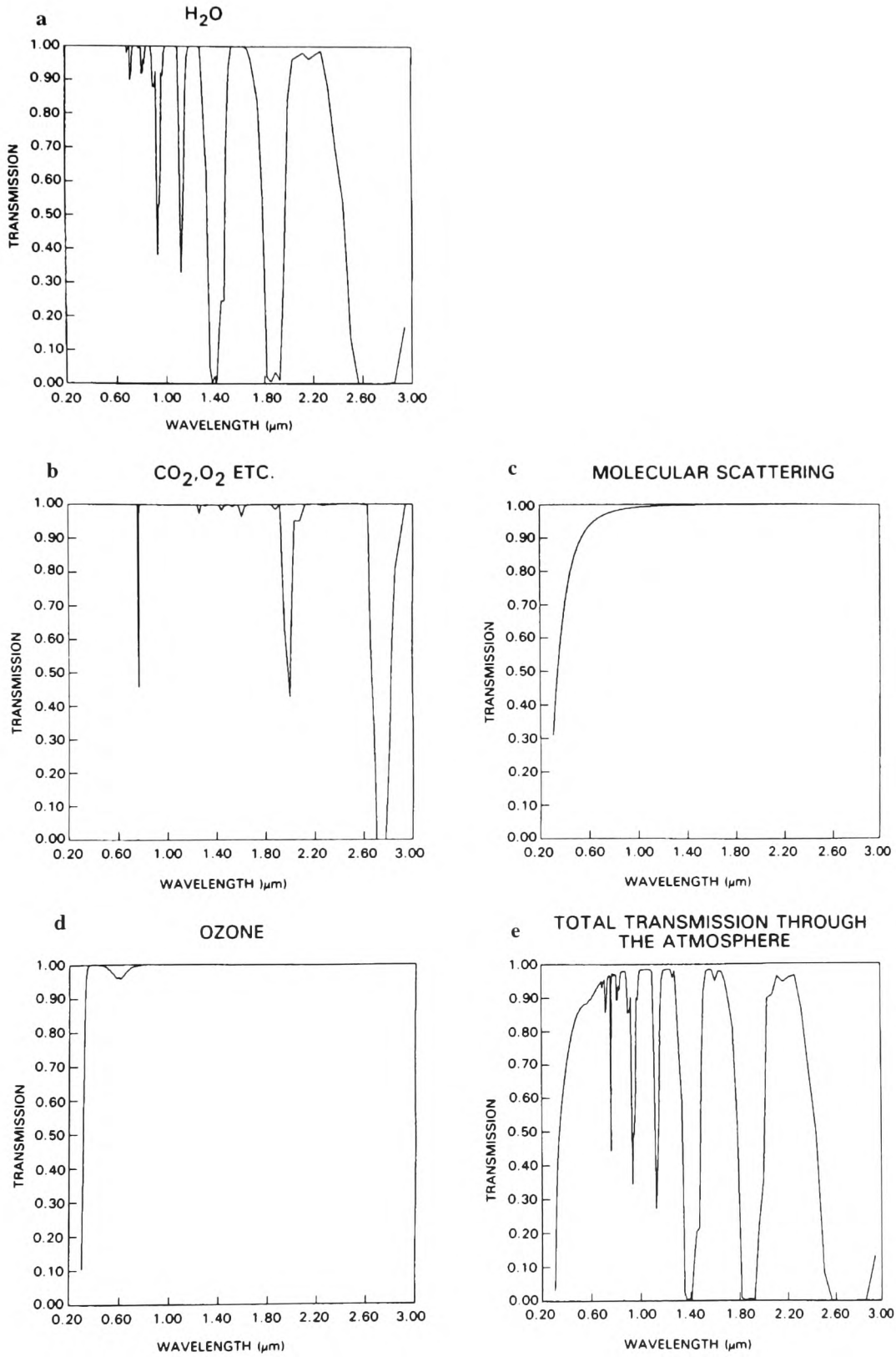


Figure 2.1 Atmospheric transmission for the sun at zenith calculated using the LOWTRAN model. Redrawn from Asrar (1989). Absorption by (a) water, (b) CO_2, O_2 and other minor absorbers, (c) molecular scattering, (d) ozone, (e) total transmission

The advent of multi-look sensors in recent years has facilitated the development of superior correction algorithms. These sensors view the same target at two or more different zenith angles relative to the sensor, a short time apart. The difference between the two signals is a result of the different optical path taken by the two soundings, and any non-isotropic scattering characteristics of the target. The development of atmospheric correction algorithms is a field in itself, and will not be covered in this thesis. All remotely sensed data are consequently assumed to have been corrected before use.

Most studies have used an empirical approach, correlating plant attributes directly with spectral reflectance. This assumes that the remotely sensed signal represents only the variable under investigation, and is independent of other parameters. This is not strictly true: for example, the presence of both water and nutrient stresses may alter the architecture of the canopy or the greenness of the leaves, which would lead to a reduction of the light intercepted and absorbed by the canopy and hence reduce photosynthesis, and since there may be more than one cause of this reduction, the remotely sensed signal does not uniquely represent the conditions being observed. Whilst the use of optical wavebands has been widespread, it has also been fraught with problems. Atmospheric effects are highly significant at these wavelengths and atmospheric water vapour in the form of clouds, omnipresent at temperate latitudes, is a serious handicap for continuous data acquisition. The use of "transmission windows" is therefore, essential, and restricts soundings to a small number of narrow wavelength ranges. The target reflectance also varies with sun angle and consequently satellite platforms tend to be in sun-synchronous orbits. Visible sensors are further restricted to daylight-only operation.

2.1.2. Vegetation Indices

Vegetation canopies are characterised by a marked contrast between a small reflectance in the visible (in particular, red) wavelengths and a large reflectance in the NIR (Figure 2.2a). In contrast, the reflectance of most soils is almost the same in both spectral regions (Stoner and Baumgardner 1981). Red light (0.63 - 0.70 μm) is strongly absorbed by chlorophyll, resulting in a small reflectance in the red part of the spectrum for green vegetation. NIR radiation is not strongly absorbed by green vegetation and suffers a high degree of scattering within the canopy. Thus proportionately more spectral radiance escapes than is absorbed, and the NIR radiance is enhanced over that of the background. Remote sensing instruments, by using combinations of reflectance data measured in the red and NIR wavebands, can exploit the difference in spectral reflectance between vegetation and soil. These spectral comparisons are known as spectral vegetation indices (SVI) or just vegetation indices, and have been widely used to infer information about vegetation from remote sensing because of their sensitivity to "greenness". Indices are devised in order to enhance the spectral contribution of green

vegetation whilst minimising the contribution from the soil background, solar irradiance and sun angle (Tucker 1979).

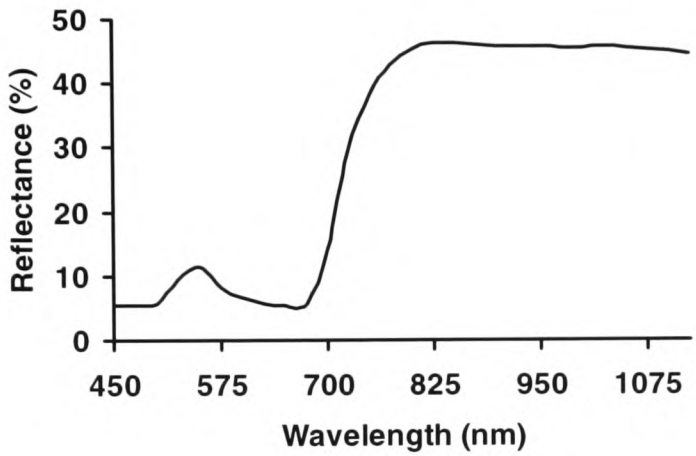


Figure 2.2a Spectral reflectance of a healthy leaf. The abrupt change in reflectance in the region between 650 and 800 nm is known as the "Red Edge". Between 400 nm and 700 nm (visible wavelengths) the reflectivity is relatively small because of absorption of radiation by leaf pigments: the maximum occurs near 500 nm (green part of the spectrum) and minima occur near 400 nm and 670 nm, caused by absorption by chlorophyll. In the NIR region, 700 nm to 1100 nm, the reflectance is relatively large, most radiation being multiply scattered by the leaf mesophyll. Multiple scattering occurs at sites where there is a discontinuity of refractive index. Within leaves, such conditions occur at the boundaries between air cavities and cell walls - see Figure 2-2b. Between 1300 nm and 2500 nm (not shown on graph), leaf reflectance is dominated by the radiation absorption of water present within the leaf tissue. Water absorption bands occur at 1430 nm and 1950 nm.

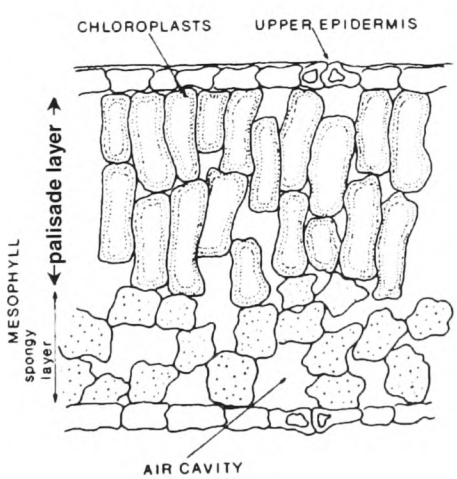


Figure 2.2b A transverse section through a leaf

More than twenty vegetation indices have been proposed, but the discussion here will be confined to the two most widely used indices, the Simple Ratio (S_R), and the normalised difference vegetation index (R_{NDVI}), This is because the applications of these two indices have far outnumbered the applications of all other indices and the advantages and disadvantages of S_R and R_{NDVI} are representative of vegetation indices as a whole. S_R and R_{NDVI} are defined as follows:

$$S_R = \frac{R_{NIR}}{R_{VIS}};$$

$$R_{NDVI} = \frac{R_{NIR} - R_{VIS}}{R_{NIR} + R_{VIS}} \qquad (2.1a,b)$$

Where: R_{NIR} is the reflectance in the near infrared,

R_{VIS} is the reflectance in the visible part of the spectrum

R_{NDVI} is the normalised difference vegetation index.

The normalisation is introduced in the NDVI in order to attempt to reduce the effects of varying solar irradiance (Jackson *et al.* 1983). Unfortunately many field studies have shown that most SVIs are sensitive to soil background reflectance and sun-view geometry, (Heute 1987; Heute 1988, Heute and Jackson 1988). Figure 2.3a shows the sensitivity of NDVI to soil brightness. Decreasing soil red reflectance causes a significant increase in NDVI. The Figure 2.3a shows that NDVI is almost as sensitive to darkening soil as it is to vegetation development. Thus, a very small amount of grass phytomass (320 kg ha^{-1}) covering a dark soil substrate had almost the same NDVI value (0.3) as a much larger amount of phytomass (1000 kg ha^{-1}) covering a bright soil substrate. Moreover, the NDVI values for a 20% cover cotton canopy, with a leaf area index of 0.5, approach those for a 60% cover, 1.7 LAI cotton crop over dark substrates. Figure 2.3b shows the NDVI of a 40% cover cotton canopy plotted against bare soil NIR reflectance for two different solar zenith angles. As the solar zenith angle increases, a larger proportion of the soil will be in shadow and a larger proportion of the vegetation will in sunlight. This results in the NDVI values being larger and a lessening of the influence of soil reflectance at larger solar zenith angles; both of these are clearly illustrated in Figure 2.3b. This also illustrates that the sensitivity of NDVI to solar zenith angle is primarily a soil effect, i.e. the sensitivity is minimal with dark soils and maximal with light soils. The dependence on soil reflectance of vegetation indices such as NDVI poses a severe limitation on their usefulness. In arid regions, where the vegetation densities are low and complex spatial patterns may exist, the noise introduced into the NDVI signal by variations is soil reflectance is likely to prove more significant than for agricultural systems, where the vegetation density is much higher and more uniform. However, variable soil moisture conditions in these systems may still pose problems.

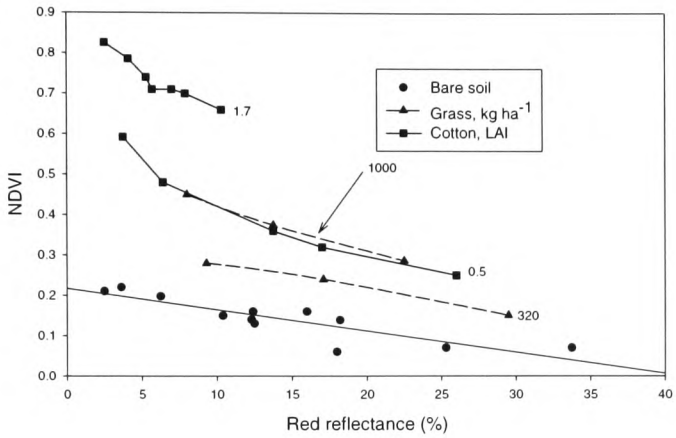


Figure 2.3a Relationship between NDVI and soil red reflectance (670 nm) for cotton and grass. The numbers next to lines indicate biomass in kg ha $^{-1}$ for grass, and LAI for cotton. Redrawn from Heute (1988).

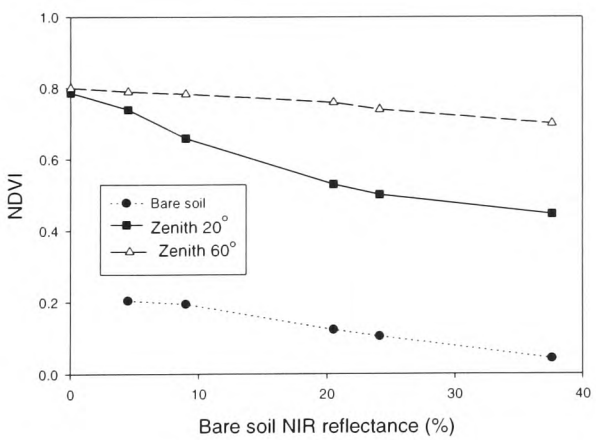


Figure 2.3b NDVI for a cotton canopy with 40% coverage, for different soil background reflectances, at two different zenith angles. Redrawn from Heute (1987).

2.1.3. Applications of Optical Remote Sensing

The are a large number and type of applications of optical remote sensing preclude (Asrar 1989) In this section I confine the discussion to a brief description, in general terms, of some of attempts at applying remote sensing techniques to deriving information which may be of use in vegetation modelling.

Over the last fifteen years there have been many attempts to form empirical relationships between SVIs and a number of vegetation canopy properties, from structural properties such as leaf area index (LAI) (Curran 1981, Asrar *et al.* 1986) and biomass (Tucker 1977, Kimes *et al.* 1981), to APAR (Kumar and Monteith 1981), evapotranspiration (Ramakrishna, Nemani and Running 1989), photosynthetic rate (Sellers 1989) and canopy biochemistry (Peterson and Running 1989). The reason why these relationships can be established is through a common link with LAI. That is, the rate of change of the SVI (S_{VI}) with respect to leaf area is proportional to the rate of change of the given property (P_p) with respect to leaf area. Mathematically this is written:

$$\frac{\partial S_{VI}}{\partial L} \propto \frac{\partial P_p}{\partial L} \tag{2.2}$$

Where: the property P_p is APAR, evapotranspiration, photosynthetic rate etc. and
L is the leaf area index

In the visible region soil reflectance is typically of the same order as canopy reflectance, so that SVIs are mainly a function of the NIR reflectance. Visible reflectance terms are included in SVIs for the simple reason that remote sensing systems measure radiances and not reflectances and consequently these measurements are a function of the incident flux. Considering this, one can see that the remote sensing instrument is really providing a result that is the product of two unknowns, the incident flux and directional reflectance of the surface. However, under normal circumstances the ratio of incident VIS and NIR radiation can be assumed to be constant, so that one is able to estimate the ratio of the surface reflectance in the VIS and NIR bands without knowledge of the incoming radiances (Sellers 1989).

From experimental results obtained by measuring the NDVI of vegetation targets it has been established that NDVI increases with leaf area index (LAI) or biomass in an asymptotic manner (Figure 2.5 and Figure 2.6). As can be seen in Figure 2.6, the NDVI *vs* LAI curve typically reaches an asymptote at a LAI of around 3, above which the NDVI signal is saturated, i.e. further increases in leaf area do not result in larger values of NDVI. This occurs because vegetation canopies are three-dimensional structures. Usually, leaves, stems or branches near the top of a canopy overlie some of those lower down. Consequently, depending upon the leaf angle distribution (LAD) and the angle of

view, NDVI increases until leaves in the upper layers completely obscure leaves underneath and the ground surface. When this occurs, then, in that direction of view, the surface appears completely green, and further increases in LAI will not result in increases in the NDVI measured. Index saturation is the main limitation of NDVI. Biomass is usually closely correlated with LAI and, therefore, exhibits a similar asymptotic relationship (Figure 2.5).

Leaf Area Index and Biomass

LAI, defined as the total one-sided leaf area per unit ground area, (Watson 1937) is one of the most important canopy structural parameters (Chason *et al.* 1990). While LAI has been used successfully in many models of both vegetation growth and fluxes of carbon dioxide and water vapour produced by vegetation, its measurement poses many difficulties. Physically harvesting leaves and other plant material is the most accurate means of determining LAI or biomass, but is extremely labour intensive, destructive and for some species, notably tall trees, may be impossible. A number of methods, both direct and indirect, have been developed for measuring LAI in the field (e.g. Ross 1981, Campbell and Norman 1989). Allometric relations between more readily measured variables and LAI have been formed, but these are species specific empirical relationships and errors are difficult to estimate. As an alternative to direct measurement, several indirect techniques have been developed using measurements of either diffuse or beam radiation penetration through the canopy (Rich 1990, Welles 1990). These approaches are designed to measure the gap fraction of the canopy, which is closely related to the penetration of solar radiation through the canopy. Given the gap fraction a model of radiation penetration is then inverted to estimate the LAI. The measurement of the gap fraction can be derived from hemispherical photographs made under diffuse lighting conditions (Madgwick and Broomfield 1969) or from an array of line sensors placed underneath the canopy (Norman and Campbell 1989). The most widely used commercial LAI measuring system, the LI-COR LAI-2000, uses this approach. The majority of gap fraction models that have been developed have assumed that, although a canopy has a specific leaf angle distribution (LAD), foliage is distributed randomly through the canopy. Unfortunately the assumption of randomness is invalid for many canopies, especially forests, and the use of these models has resulted in a large underestimation of LAI being reported for several forest stands, e.g. Chason *et al.* (1990) for oak-hickory, and Gower and Norman (1991) for oak-spruce.

Lang and Xiang (1986) introduced an alternative gap fraction modelling approach, in which measurements of beam transmission along a transect are broken down into sub-groups, the average value for each sub-group is calculated, and the sub-groups are combined by averaging the logs of the sub-group values. Such an averaging procedure accommodates natural discontinuities in the canopy,

such as row effects. Good results have been reported using this technique (e.g. Fassnach *et al.* 1994), and the approach is used in the commercially available DEMON system (Lang *et al.* 1985).

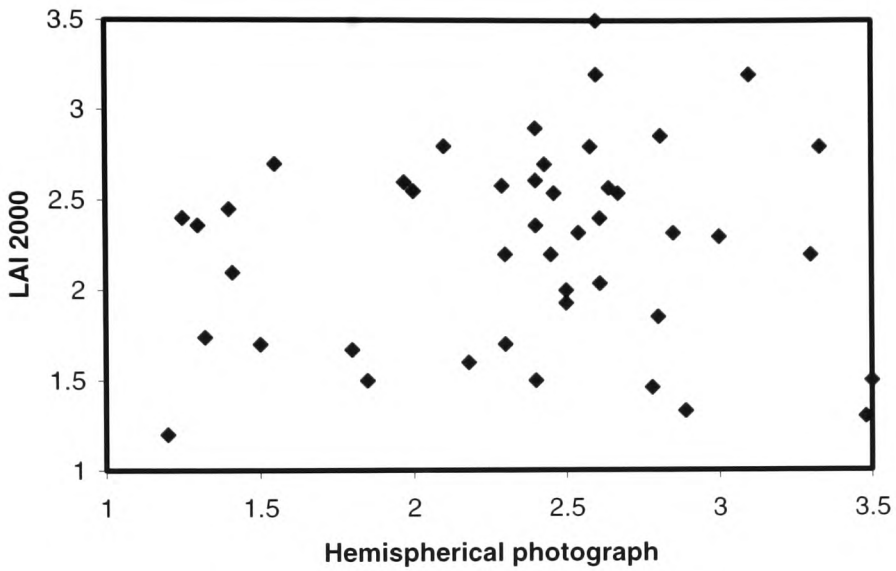


Figure 2.4 Leaf area index measured by two different ground-based methods. Redrawn from Ross and Nilson (1983)

The technique of taking hemispherical photographs has also been widely used, and is analogous to the use of the LAI-2000. A camera is mounted below the canopy, facing upward. Photographs are taken using a fish-eye lens and then analysed to estimate the LAI. All three methods, LAI-2000 hemispherical photography and DEMON, use an algorithm that assumes that the foliage is randomly distributed in space. This may introduce significant errors for clumped canopies such as forests. The hemispherical photography and DEMON methods both enable the image to be segmented, so that area of non-random foliage distribution may be excluded, improving the accuracy of these techniques.

Whilst indirect methods for the field measurement of LAI are extremely useful, it is vital that the models of canopy architecture on which they are based are realistic descriptions of the canopy structure. The use of simplifying assumptions in these models can lead to large errors in the LAI estimates obtained. In practice, if two or more indirect techniques are employed to measure the LAI of the same canopy stand, the results may vary significantly (Figure 2.4). Such differences may result from spatial inhomogeneities of the canopy, or problems with instrumentation, for example, setting the exposure for hemispherical photographs, or obtaining readings with perfectly diffuse incident radiation with the LAI-2000. The problems involved in estimating LAI from remotely sensed data must be weighed against the problems of making accurate estimates on the ground. Field

measurements of LAI may carry an error of, at best, 10%, therefore errors of this magnitude from remotely sensed measurements must be regarded as more than acceptable.

In addition to the leaf area, the leaf angle distribution (LAD) affects canopy reflectance through the changing geometry of the sun-canopy-sensor system. Canopies with mostly horizontal leaves have the least variability in reflectance with changes in solar and view angles. For those with mostly erect leaves as the solar zenith angle increases, reflectance decreases in the visible region, but increases in the infrared region, and as the view angle increases so does the reflectance in both wavebands. Owing to this complicated behaviour leaf angle distributions only tend to be considered for canopy radiation transfer models.

Absorbed Radiation

Photosynthesis is stimulated by light in the visible part of the spectrum between approximately 400 nm and 700 nm. This waveband is commonly referred to as photosynthetically active radiation (PAR) (Monteith and Unsworth 1990). In early studies, PAR was applied to radiation measured in units of energy flux density (W m^{-2}). The net absorbed energy in all wavebands is the driving variable for transpiration but for some processes the effect of radiation is more dependent on the number of photons absorbed than on their energy. Photosynthesis is one such process and this case it is more appropriate to express radiation in term of the photosynthetic photon flux density or PPFD.

The amount of energy or photons absorbed (I_{APAR}) is determined by the fraction of incident radiation which is reflected and transmitted by the canopy elements, and may be written as:

$$I_{APAR} = (I_i + I_s) - (I_c + I_t) \tag{2.3}$$

Where I_i is the incident flux

I_s and I_c are the reflected energy (or photon flux) from the soil and plant canopy

I_t is the flux transmitted through the canopy.

The ability to estimate I_{APAR} (or APAR) from remote sensing depends upon the contrast between reflectances of soil and vegetation. Moreover, remote sensing platforms can only measure one of the four components in the equation above, the reflection. To overcome these difficulties, either complementary relationships at different wavelengths (i.e. vegetation indices), or models representing the reflectance of the individual surface components (vegetation and background), are used.

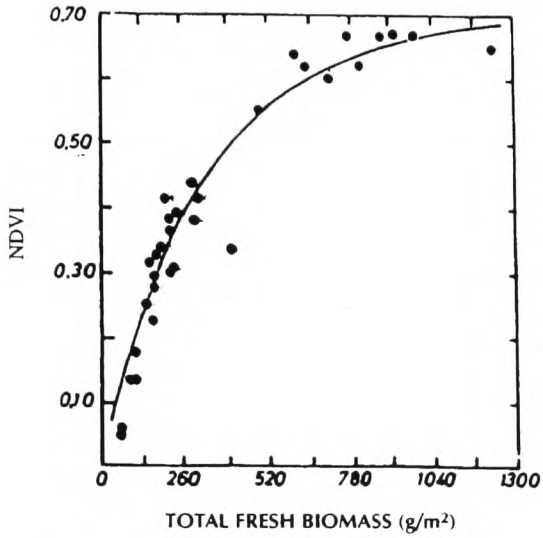


Figure 2.5 An example of the relationship between NDVI and biomass, according to Tucker (1979)

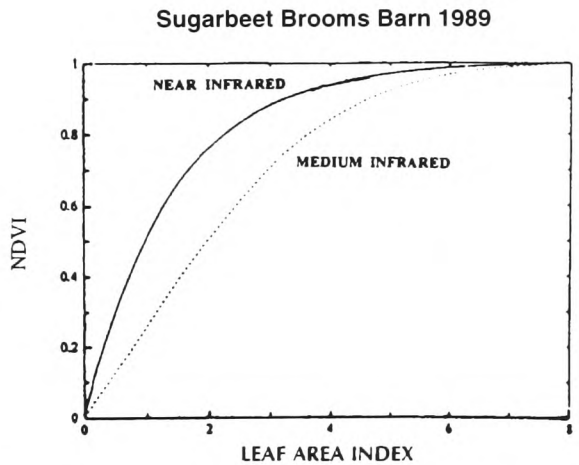


Figure 2.6 The relation between two vegetation indices and LAI. NDVI is represented by the solid line. The dotted line represents a short-wave infrared index as proposed for the "Vegetation" instrument.

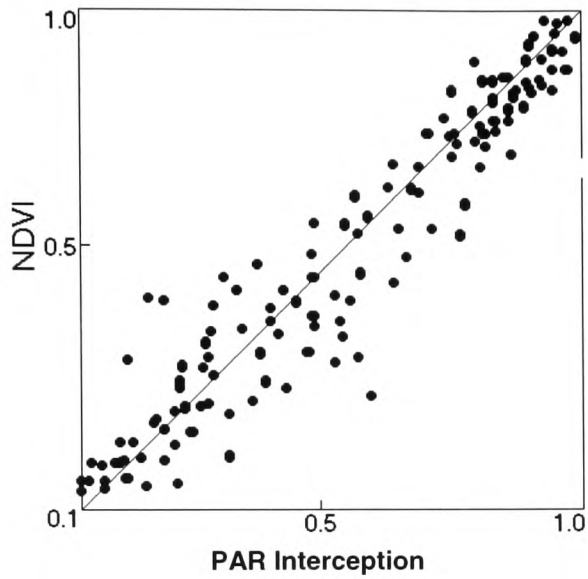


Figure 2.7 The relationship between NDVI and intercepted PAR for a number of different wheat canopies. Redrawn from Asrar *et al.* (1984)

Figure 2.7 shows the relationship between NDVI and intercepted PAR for several wheat canopies. The overall relationship is quite good, because the surfaces that reflect radiation to the sensor are those that are exposed to solar radiation. However, during the early part of the growing season, identified by small values for NDVI, the scatter of data points is larger than later in the season. In early stages of growth the canopy is not closed and the background (soil) surface is exposed to view. The sensor measures the combined reflectance from both the vegetation and the background. Vegetation indices such as NDVI are sensitive to changes in background reflectance caused by rainfall or irrigation, resulting in the scatter seen in Figure 2.7. Atmospheric conditions and solar zenith angle also change during the growing season that may also explain some of the observed scatter. For closed canopies SVIs can give good estimates of APAR. However, for open canopies models of surface reflectance are required, which when combined with remotely sensed data can give estimates of APAR.

Vegetation stress

A major factor limiting plant growth, especially for natural as opposed to agricultural systems, is the lack of water. When insufficient water is available to meet the transpirational and physiological demands of a plant, water stress may develop. When vegetation is under stress, problems occur with the use of remotely sensed parameters, especially with productivity estimates. The fraction of sunlight

absorbed by the canopy has been shown to be the main determinant of productivity (Monteith 1977), and can be expressed as:

$$\frac{dW}{dt} = efs \tag{2.4}$$

Where: W is dry mass
s is incident PAR
f is the fraction of PAR absorbed by the canopy
e is the conversion quotient

Effects of stress on vegetation can be manifested as a change in leaf area or PAR interception, partitioning of biomass between leaves, roots, seeds etc., or in the efficiency of energy conversion. The latter may be the result of chlorosis (loss of greenness in leaves), closure of stomata, or changes in internal CO_2 metabolism.

Both near-infrared (NIR) and thermal-infrared wavelengths have been used for the detection of water stress in plants. In the NIR changes in plant reflectance caused by water stress result from changes in internal leaf structure causing an increase in scattering from the leaf (Boyer *et al.* 1988). Reductions in efficiency, or variations in stomatal conductance caused by, for example, water stress are difficult to estimate with remote sensing, whereas changes in leaf area or chlorosis have been measured with some success. Riggs and Running (1991) measured the reflectance of water-stressed Norway spruce (*Picea abies*) and white pine (*Pinus strobus*) in Germany using the high spectral resolution Airborne Imaging Spectrometer (AIS). They found that differences in the reflectance between stressed and unstressed plots could only be detected when severe water stress was present, and when the plots were imaged simultaneously. Differences in reflectance decreased to insignificant levels when the sites were imaged three hours apart. Moreover, the study was carefully designed to minimise the effects of sun-angle, look-angle, atmospheric variability and surface heterogeneity. Under operational remote sensing conditions, the detection of water stress using optical methods is probably impossible (Pierce, Riggs and Running 1990, Hunt and Rock 1989). For trees only very severe water stress leading to defoliation (-3 to −5 MPa) would appear to be detectable, and in conditions this severe the plants would be probably dead within a few weeks making detection easy using a number of optical sensors.

Several water stress studies have used thermal-infrared remotely sensed surface temperatures, and an energy budget logic which assumes that plants respond to water stress through stomatal closure, thereby decreasing the latent heat flux and increasing leaf temperature (e.g. Pierce and Congalton 1988). Pierce, Riggs and Running (1990) used both thermal, visible and infrared data in an attempt to identify water stress in Norway spruce and white pine. They concluded that thermal techniques were no better than other optical techniques for determining water stress in conifers, because of the extreme stress needed before detection is possible.

Very High Spectral Resolution Indices

The spectral width of the channels used to compute the NDVI from Landsat is relatively large. Studies have been made to determine the usefulness of very narrow wavelength channels for the formation of very high spectral resolution indices. New remotely sensed information is being received from current developments in sensor technology, typified by platforms such as AIS (Airborne Imaging Spectrometer), AVRIS (Airborne, Visible, Infrared, Imaging Spectrometer) and MODIS (Moderate Resolution Imaging Spectrometer). Studies of the observed spectral features at leaf or canopy scale have tended to concentrate on the 'red edge' where optical properties change abruptly with wavelength (Figure 2.2). The spectral features of the red edge have generally been described by studying either the first or second derivatives of spectra with respect to wavelength, or spectral shifts of the inflection point of the red edge (Hall *et al.* 1990, Curran et al 1991).

So called 'broad band' vegetation indices such as NDVI have been used extensively and are well correlated to LAI and APAR, but they are susceptible to soil background interference, and suffer from signal saturation for LAI > 3. The good correlation between SVIs and absorbed radiation, combined with the widespread availability of SVI data sets, makes their use for vegetation modelling attractive. As bi-directional reflectance measurements (BRDF) become more readily available from new, high spatial and spectral resolution instruments, they are likely to replace SVIs because of the larger information content within the BRDF signature. The use of BRDF data will, however, require the further development of models to represent the anisotropic scattering behaviour of vegetation canopies.

Thermal Infrared Remote Sensing

Remote sensing measurements in the thermal infrared are more complicated than measurements in the visible or near infrared wavebands, for the following reasons.

- Surface elements are capable of storing and then releasing heat later.
- Heat energy is gained and lost not only through radiative processes, but also through exchanges of sensible and latent heat.

The radiation that is absorbed at the Earth's surface (called the net radiation, R_n) is the difference between incoming radiation in all wavelengths, and the reflected shortwave (0.15 – 4 μm) and both the reflected and emitted longwave radiation (> 4 μm)

The net radiant energy absorbed by the surface is dissipated through conduction into the surface (usually called soil heat flux, G), convection into the atmosphere, H, as latent heat of vaporisation of water, E, and to a lesser extent by photosynthetic processes, i.e.

$$R_N = G + \lambda E + H \tag{2.5}$$

Where R_N is the net radiation flux density

G is the soil heat flux.

λ is the latent heat of vaporisation of water.

E is the evaporation flux of water

H is the sensible heat flux.

The thermal infrared spectral region covers the range from approximately 3 to 100 μm, but only the spectral *windows* between 3.4 to 4 μm and 8 to 14 μm are of interest for remote sensing (Figure 2.1). Outside these windows, the atmosphere is almost opaque to radiation. This opacity is less for low-flying aircraft than for satellite sensors since the former observe the surface through a much smaller depth of atmosphere. The opaque regions are not wholly without interest, however: satellite measurements of radiation at these wavelengths provide information about the temperature and humidity profile between the surface and the atmosphere.

Net radiation can be expressed as the sum of four components:

$$R_n = R_{sd} - R_{su} + R_{ld} - R_{lu} \tag{2.6}$$

Where: R_{sd} is downwelling shortwave radiation

R_{su} is the shortwave radiation reflected upwards by the surface

R_{ld} is the longwave radiation emitted by the atmosphere toward the surface

R_{lu} is the longwave radiation emitted by the surface into the atmosphere

Measurements of R_{sd} are straightforward to obtain using a pyranometer, and satellites can easily measure the reflected component R_{su}. However multi-spectral radiometers typically only measure a part of the total reflected energy, and consequently the fractional contribution that the sensor measures for each waveband must be multiplied by an appropriate conversion factor. The term R_{ld} is given by:

$$R_{ld} = \varepsilon_a \sigma T_a^4 \tag{2.7}$$

Where: ε_a is the atmospheric emissivity

σ is the Stefan-Boltzmann constant

T_a is the air temperature at screen height

Various models exist for the determination of ε_a but these require knowledge of the water vapour content of the atmosphere. In addition T_a is required from local, ground based, measurements or needs to be inferred from an atmospheric model such as LOWTRAN. The longwave radiation emitted by the surface is given by an equation analogous to (2.7), i.e.

$$R_{lu} = \varepsilon_s \sigma T_s^4 \tag{2.8}$$

Where: ε_s is the surface emissivity

T_s is the surface temperature

Unfortunately, ε_a and T_s are frequently unknown, and setting $T_s = T_a$ is likely to introduce serious errors. For example, bare soil could be 25 °C hotter than T_a or a well-irrigated closed canopy 10 °C cooler.

Thermal remote sensing and the radiation balance equation (2.6) have been used to derive net radiation, evapotranspiration and surface temperature. Unfortunately, because of the difficulty in making atmospheric corrections, results have often been poor, with errors typically being of the order of 20% for R_n (Asrar 1989), 30% for T_s (Kiang 1982). Estimation of evapotranspiration is essentially a matter of estimating the latent heat term (LE) for the energy balance equation (Running 1991, Ramakrishna, Nemani and Running 1989). However, even under very favourable conditions (aircraft measurements at 150 m, and a homogeneous, complete, canopy cover) significant errors can occur, as shown in Figure 2.8 (Jackson *et al.* 1987). These errors are commonly because of inaccurate determination of the sensible heat flux, H, which is commonly derived from:

$$H = \frac{(T_s - T_a)\rho C_p}{r_{ah}} \tag{2.9}$$

where: T_a is the temperature at screen height

T_s is the surface temperature

ρC_p is the volumetric heat capacity of air

r_{ah} is the convective resistance to heat transfer

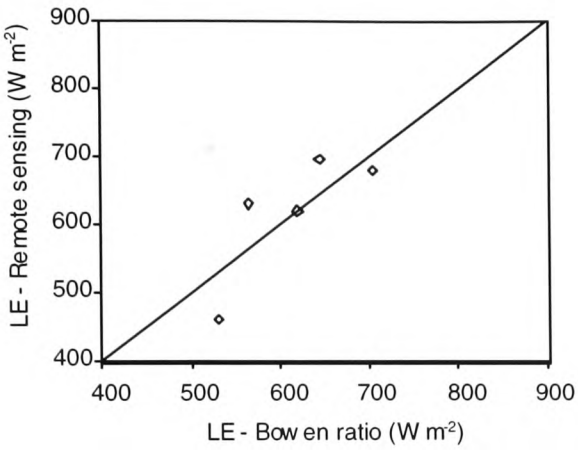


Figure 2.8 Instantaneous latent heat flux estimated by remote sensing and ground-based Bowen ratio measurements. Redrawn from Jackson *et al* (1987)

Errors in estimation of the variables in equation 2.9 increase with increasing wind speed and increasing canopy roughness and the effects of any errors on the estimation of latent heat flux become more pronounced when $H \approx LE$. At satellite altitudes, the much longer path between the surface and the sensor contains a much larger amount of water vapour, increasing the uncertainty in the measured longwave radiation. The effects of uncertainties in the surface emissivity in equation 2.8 are associated with errors of approximately ±0.07 °C for ±0.01 ε_s (Huband and Monteith 1986). Atmospheric effects, combined with other uncertainties lead to inaccuracies in T_s derived from satellite measurements of between 2 and 4 °C under most conditions (Asrar 1989). All of these difficulties are exacerbated by the spatial inhomogeneties of vegetation surfaces, necessitating the use of spatial averaging schemes to determine surface emissivity, temperature and roughness. Moreover, remote sensing offers only instantaneous measurements of parameters such as evapotranspiration: detailed meteorological data and models are required to extrapolate these point data to daily values.

The estimation of soil moisture using visible and near infrared remote sensing techniques has not proved very successful because the spectral radiance at these wavelengths depends upon several other variables, e.g. surface roughness, geometry of illumination, organic matter content and soil texture. Soil moisture studies have usually been based on measurements of soil albedo, which has been found to be linearly related to soil moisture. Similar relationships have been shown for deeper layers, because of the correlation between deep and surface moisture, but this technique has two distinct drawbacks: (1) correlation for deeper layers breaks down with small overall water content, and (ii) universal relationships are difficult, if not impossible, to achieve because of the wide variation in the albedo of different soils.

The general equation for change in soil moisture content, θ, can be written:

$$\frac{d\theta}{dt} = P_{pt} - R_f + L_f + E_{vap} - T_r + C_{pr} - Q_p \qquad (2.10)$$

Where: P_{pt} is precipitation,
E_{vap} is evaporation,
T_r is transpiration,
R_f is surface runoff,
L_f is lateral sub-surface flow,
Q_p is percolation
C_{pr} is capillary rise from lower levels.

Thermal remote sensing techniques for measuring soil moisture are based on the concept of thermal inertia (P_c), which is defined as:

$$P_c = \sqrt{K_{soil} C_{soil}} \qquad (2.11)$$

Where: K_{soil} is the thermal conductivity of the soil
C_{soil} is the specific heat capacity of the soil

Thermal inertia is a measure of the soil internal resistance to temperature change. Both C_{soil} and K_{soil} increase with increasing soil moisture, and thus so does the thermal inertia. Therefore, the soil water content may be inferred from a time series of soil temperature measurements. Soil surface temperatures are influenced by deep water content, by external variables, such as solar radiation, air temperature, relative humidity, cloudiness and wind, and by internal variables, principally C_{soil} and K_{soil}, which vary with soil type. A complicating factor is evaporation, which reduces the amplitude of the diurnal surface temperature cycle. As a result, the day-night temperature difference is an indicator of some combination of soil moisture and surface evaporation. In addition the presence of a vegetation overstory may wholly, or partly, obscure the soil from view, restricting the use of optical and thermal techniques for measuring soil moisture to sparsely vegetated areas.

The quantitative interpretation of remotely sensed thermal data is complicated by the many physical factors that influence the observations. These include the following:

- Evapotranspiration which greatly influences both daily average surface temperature and day to night variance.

- Water vapour in the atmosphere, which modifies downwelling visible and upwelling thermal radiation, whilst also modifying the satellite observed radiances. Accurate atmospheric corrections are therefore essential.

- The air temperature near the ground, which affects the exchange of sensible heat between the surface and atmosphere. This is a very important term, but can usually be estimated from conventional, local, meteorological data sources. In the absence of such data an atmospheric prediction model is required.

- Surface roughness and wind speed, which affect the transfers of sensible and latent heat. In addition atmospheric humidity influences the magnitude of evapotranspiration through stomatal closure.

- The heat-storing capacity of soil, which affects the day-night temperature changes. This factor is more important in arid areas where there is little evaporation.

- Vegetation surface emissivities are often poorly known.

- Topography and the sheltering effects of other canopies, which modify the heat balance of vegetation.

From this list, it is not surprising that thermal remote sensing has not achieved widespread and regular use in the remote sensing community.

2.1.4. Conclusions

Over the past twenty years, studies with VIS and NIR remote sensing have far outnumbered those using other wavebands. This is understandable, both because of the far greater number of VIS and NIR sensing platforms that have been built, and because of the ability to exploit the contrasting reflectance properties of vegetation in these two wavebands. Applications for vegetation monitoring, whilst enormously varied in terms of the properties involved, have concentrated mainly on the construction of empirical relationships between these properties and vegetation indices (VIs). These relationships may be established providing the rate of change of the VI with respect to leaf area is proportional to the rate of change of the given property with leaf area. Unfortunately, VIs suffer from signal saturation at relative small LAIs and are sensitive to changes in background reflectance, caused for example, by rainfall.

Early sensors were single look instruments (i.e. they observed the surface at one zenith angle, usually close to nadir), however the reflectance of light from vegetation canopies is highly anisotropic and characterised by the bidirectional reflectance distribution function (BRDF) of the canopy. Recently

multi-look sensors have been developed that allow the BRDF to be reconstructed from reflectance measurements. Multi-look measurements also facilitate atmospheric corrections by observing the same target through different atmospheric paths.

BRDF data contain much more information about the surface than VI measurements and therefore are the way forward in optical wavelength remote sensing. Although active microwave techniques offer some advantages over optical techniques (e.g. insensitivity to cloud cover), microwave remote sensing datasets will be difficult to obtain until new radar remote sensing platforms have been developed. This means that BRDF remote sensing datasets are currently the best choice for many applications. Nevertheless, the SVAT parameters obtainable from BRDF data are limited to LAI, biomass and APAR. In order to interpret multi-look data, BRDF models are required: these are considered in Chapter 3.

2.2. MICROWAVE WAVELENGTHS

Microwave techniques can be broken down into passive systems, which detect the microwave thermal emissions from the earth's surface, and active systems, or radars, which transmit pulses of microwave radiation towards the earth and collect the reflections. Radars offer very high spatial resolution (up to 4 m) but involve highly complex, post-acquisition processing; the returned backscatter signal can also be difficult to interpret. Passive sensors collect the very weak emissions from the surface and

P Band	*L Band*	*C Band*	*X Band*
0.225 - 0.39 GHz	0.39 – 1.55 GHz	4.2 – 5.75 GHz	5.75 – 10.9 GHz
AIRSAR	AIRSAR, SIR-C, RADARSAT	ERS1, RADARSAT, SIR-C	XSAR
Good estimates of standing biomass	Separates broad leaf and fallow sites	Good discrimination of crop types	Intermediate discrimination properties
Sensitive to row direction	Sensitive to row direction	Insensitive to row direction	Insensitive to row direction
Very deep penetration of canopy	Deep penetration of canopy	Shallow penetration of canopy	Does not penetrate canopy

Table 1: Microwave waveband classifications and properties for remote sensing

consequently have poor spatial resolution.

2.2.1. Passive Microwaves

A passive microwave radiometer measures the thermal emission from the Earth's surface. The strength of passive microwave radiation largely depends upon the temperature and dielectric properties of the surface. The dielectric property of a material is characterised by its dielectric constant, which is a measure of the propagation and energy loss behaviour of electromagnetic waves in the material. The large value of the dielectric constant for water of about 80 (at 20 °C and 1GHz) contrasts with values for other natural materials of 3 to 8 under the same conditions, thus making passive microwave remote sensing techniques useful for studying water resources. Unfortunately, the small amplitude of microwave emissions from most surfaces precludes the use of this technique at sub-regional scales.

Passive microwave radiometers have been providing information about atmospheric and oceanic parameters for a number of years. However, this has not been the case for land surfaces; firstly, because the spatial resolution of passive sensors has been large relative to the scale of spatial variations of the land surface and secondly, because the microwave emission characteristics of land targets are not well understood since land surfaces tend to have complex geometric and dielectric properties. Nevertheless significant research has been undertaken with passive microwave sensors, particularly for the estimation of soil moisture.

According to the Rayleigh-Jeans approximation (Ulaby *et al.* 1982), intensity of microwave radiation emitted from a vegetated surface at any polarisation, *p*, and zenith angle, θ, to the radiometer can be expressed in terms of the brightness temperature, $T_B^P(\theta)$, as follows (Choudhury 1991):

$$T_B^P(\theta) = T_S[1 - R_P(\theta)\exp(\frac{-2\tau}{\mu})] \qquad (2.12)$$

Where: $\mu = \cos(\theta)$

T_s is the surface temperature.

τ is the optical thickness of the canopy.

$R_p(\theta)$ is the reflectivity at the soil-air interface, which depends upon soil moisture.

The above equation is valid for an extensive, homogeneous medium, but, at the spatial resoultions (100 km) of satellite passive microwave sounding, vegetation surfaces tend to be relatively heterogeneous. For these conditions one can write an approximate equation for the brightness temperature as follows (Choudhury 1991):

$$T_B^P(\theta) = T_s[1-R_p(\theta)](1-f)^{\eta+1} + T_s[1-R_p(\theta)\exp(\frac{-2\tau}{\mu})][1-(1-f)^{\eta+1}] \quad (2.13)$$

Where: $\eta = \tan\frac{\theta}{b}$

b is a dimensionless parameter governed by the geometry of the vegetation clumps.

f is the fractional vegetation cover.

Retrieval of Soil Moisture

Soil moisture content, roughness and texture exert a significant influence on the microwave backscatter signal. The significant electromagnetic characteristic that allows microwave remote sensing of soil moisture is the strong dependence of the soil dielectric constant on the moisture content. The soil moisture information contained in $T_B^P(\theta)$ is modified by the effect of soil roughness on the microwave signal. As soil roughness increases, the reflectivity decreases, and hence the brightness temperature increases. The effect of surface roughness on $T_B^P(\theta)$ is wavelength dependent and is smaller for long wavelengths than for short wavelengths (Choudhury *et al.* 1979). The vegetation cover also modifies $T_B^P(\theta)$ by attenuating the emission from the soil and by emitting radiation itself.

The optical thickness of a canopy, τ, can be defined as:

$$\tau = \frac{cW}{\lambda} \quad (2.14)$$

Where: W is the water content of the vegetation.
c is a constant.
λ is the wavelength of the radiation.

The dielectric properties of moist soil may be characterised by a frequency dependent complex dielectric response function, $\xi(\omega)$ as follows:

$$\xi(\omega) = \xi_r(\omega) + i\xi_i(\omega) \quad (2.15)$$

Where $\xi_r(\omega)$ is the real part,
$\xi_i(\omega)$ is the imaginary part and,
ω is the angular frequency.

The function ξ_r is a measure of the energy stored by dipoles aligned in the magnetic field and is almost constant up to the relaxation frequency, ω_r, while ξ_i is a measure of the energy dissipation rate of the medium. The time, $1/\omega_r$ is the time constant for the decay in polarisation when the field is

removed. When the frequency is higher than ω_r the dipoles can no longer follow the field and the ability of the medium to store electrical energy decreases - hence ξ_r decreases. In complicated heterogeneous media, there may be more than one relaxation mechanism and more than one absorption peak. In a soil, the values of ξ_r are typically between 3 and 5, whereas ξ_r for water is 80; hence extremely small amounts of water greatly affect the electromagnetic properties. At low moisture contents ξ_r increases slowly. However, above a certain point, the transition moisture content, the slope increases sharply because of a change in the behaviour of the soil; the transition moisture content depends upon soil texture and is less for sand than for clay.

The amount of water held within soil is strongly associated with the soil type and can be either *bound* or *free*. Bound water is held tightly to the soil particles by ionic forces; free water on the other hand is at a sufficient distance from surfaces to allow relatively free movement. Thus the soil particle size and distribution are important: C band microwaves with HH (see later for description of notation HH) polarisation and incidence angle of between 10 and 20 degrees seems to give the best results for determining soil moisture (Beaudoin *et al.* 1990).

Combining soil roughness and vegetation effects, longer wavelengths are better for soil moisture estimation when atmospheric effects can be ignored. Conversely higher frequency microwave signals are dominated by the vegetation signal. However, at high frequencies, Equation 2.12 no longer holds because the size of the canopy elements tends to be larger than the wavelength and consequently one must consider both scattering and absorption of radiation propagating through, and emitted by, the canopy.

Retrieval of Canopy Structure Parameters

The microwave emission signal of vegetation canopies may allow retrieval of structural parameters and micro-meteorological variables, e.g. leaf temperature and stomatal resistance (Calvet *et al.* 1992). However, limitations are the large number of vegetation parameters that influence the signal, the strong contribution to the signal from the soil background, and the coarse resolution of typical sensors, e.g. NIMBUS 7 Scanning Multi-channel Microwave Radiometer (SMMR). To combat these limitations, much of the work in this field has been carried out over the Amazon. Although human (e.g. logging) and natural (e.g. windthrow) disturbances now affect almost every area of the Amazon forest, and seasonality with respect to forest water status is evident, areas can be identified that are relatively homogeneous at large scales comparable with the resolution of the sensor (Calvet *et al.* 1992).

Numerical models of vegetation emission are necessary to interpret soundings; these generally form two classes:

1. discrete models, based on the scattering and absorbing properties of each bio-element (branches are modelled as cylinders, leaves as discs etc.), and
2. continuous models, based on a statistical description of the dielectric properties of the canopy.

For discrete models, the dimensions, orientations and dielectric properties of each bio-element need to be evaluated in the computation of the scattering and extinction functions. Such a detailed description is difficult to formulate and the burden of processing time is excessive.

In continuous models the canopy is statistically described and characterised through spatial fluctuations in ε, the permitivity of the canopy. The correlation function of ε depends upon two geometric parameters, the horizontal and vertical correlation lengths.

$$\varepsilon(r) = \varepsilon_m + \varepsilon_f(r) \tag{2.16}$$

Where : r is position.
ε_m is the spatially averaged permitivity, and depends on the surface dielectric constant and structure.
$\varepsilon_f(r)$ is the fluctuation in permitivity at position r, and is called the correlation function; it depends on two parameters, the horizontal and vertical correlation lengths.

The discrete approach is useful to investigate the sensitivity of microwave emission to vegetation and soil parameters, but its use for inversion is limited by the requirement for a precise description of vegetation structure. The continuous approach reduces the structural information to two parameters, the correlation lengths. However these are fitted parameters whose geometric and dielectric significance is hard to interpret.

In a typical example of continuous modelling retrievals of soil moisture and vegetation volume fraction, correlation lengths are calibrated over a well-developed canopy by forward modelling with ground-based data. The correlation lengths are assumed to be constant over the growing period, and the model is then inverted to retrieve the required parameters (Wigneron, Kerr *et al.* 1992). Unfortunately the assumption that correlation lengths remain constant over the growing period cannot be justified until a better understanding of their temporal evolution is achieved; this remains a major difficulty for continuous modelling.

Conclusion

Practical applications of passive microwave remote sensing for stand scale vegetation modelling are severely limited by the spatial resolution of sensors, especially space-borne instruments. Unlike visible and infrared remote sensing, microwave sensors allow soundings irrespective of cloud cover, provided there is no significant precipitation occurring, and consequently may offer useful phenological information provided a repeat period of three to four days is provided by the satellite orbit. Passive microwave estimations of soil moisture require longer wavelength observations in order to minimise the effects of surface roughness and vegetation overstory. The effective soil layer thickness, the moisture content of which determines the reflectivity, also increases with wavelength observations and is very roughly equal to $\lambda/10$. Soil moisture at deeper levels may be estimated by coupling remotely sensed brightness temperature with soil moisture simulation models (Camilo and Schmugge 1983).

Passive microwave remote sensing of vegetation structural properties and surface temperature (and hence radiation budget and evapotranspiration) is limited by the poor spatial resolution and assumed homogeneity of the surface. Complex and invertable models of microwave emission are needed for these techniques, for which parameterisation is difficult. Consequently, it is unlikely that passive microwave remote sensing will offer information useful to stand-scale ecological modellers in the near future, and is likely to be limited to providing only regional scale estimates of soil and vegetation water content. Moreover the availability of such datasets is likely to be restricted because of a lack of satellite soundings, until new sensor platforms are commissioned.

2.2.2. Active Microwaves (RADAR)

Active microwave sensors, or radars, transmit a pulse of microwave radiation toward the surface of the earth and then collect the scattered return signal. The intensity of a given pixel in a radar image is related to the backscatter coefficient, σ^o, of a surface, which can then be related to the physical properties of the surface. This relationship depends upon the microwave frequency, the angle of incidence and the receive-transmit polarisation configuration (HV meaning horizontal receive, vertical transmit etc.). For spatially periodic surfaces, e.g. crops planted in rows, the azimuth angle of the target is also highly significant. Radars offer the twin advantages of 24 hour operation and insensitivity to cloud cover. However, the interaction between microwaves and soil or vegetation is still poorly understood. Through careful design of the remote sensing platform as described below, the spatial resolution of radar systems can be quite high (less than 10 m). Although very high spatial resolution is necessary if remotely sensed data are to be incorporated into ecological modelling at stand scales, a major difficulty of using radar techniques lies in our ability to interpret the data. This

requires an understanding of both the theory of radiation transfer, and the complex design of remote sensing radar systems.

The angular resolution of a satellite instrument is defined by :

$$\theta = 1.22\frac{\lambda}{D} \tag{2.17}$$

where: θ is the angular resolution (radians)
λ is the wavelength of the radiation
D is the aperture of the instrument

For remote sensing purposes it is often convenient for the image to be in the form of a strip. This is achieved by pointing the radar sideways from the aircraft (or spacecraft), and transmitting in the direction normal to the platform motion, called the range direction (Y in Figures 2.9 and 2.10). Scanning in the azimuth direction (X in Figure 2.9) is achieved over a period of time as the aircraft moves. The resolution in the "range direction" is limited by the duration of the radar pulse, since the same antenna is normally used to both transmit and receive pulses. The resolution in the azimuth direction is limited by the width of the beam.

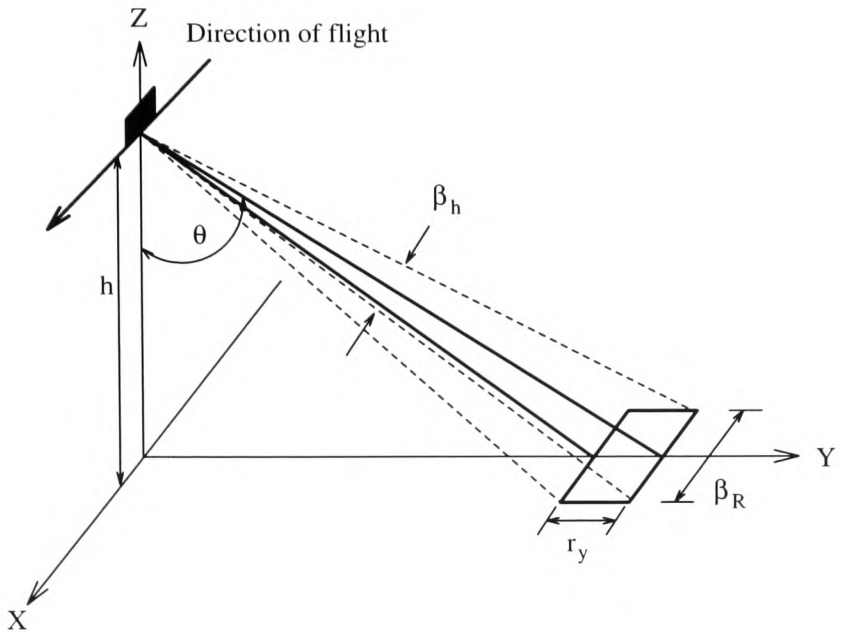


Figure 2.9: Geometry of real aperture radar - view in the *X-Y* plane.

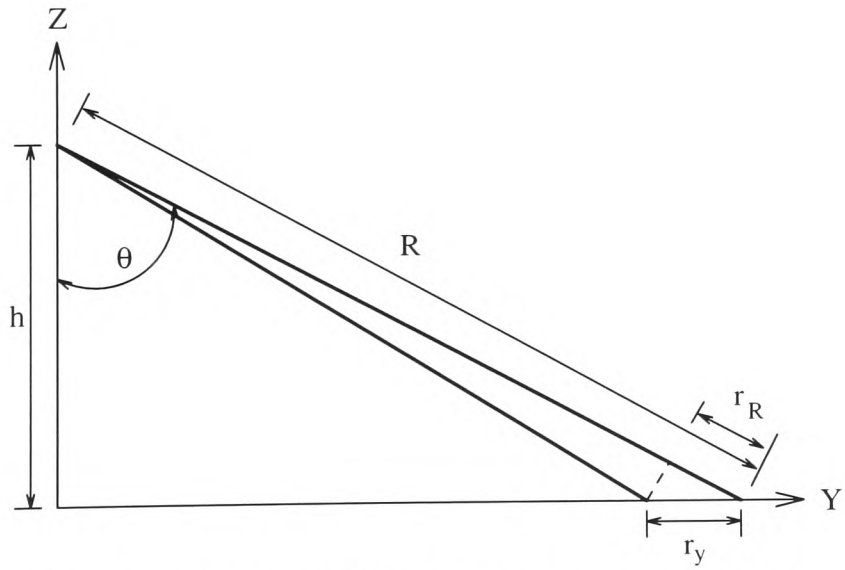


Figure 2.10: Geometry of a real aperture radar - view in the *Y-Z* plane.

The time interval between returned signals must be longer than the duration of a single radar pulse (τ_p) so that consecutive signals do not interfere with one another. The round trip time (T) for a pulse going to a point at slant range, R, is given by

$$T = \frac{2R}{c} \tag{2.18}$$

Where c is the speed of light.

Therefore the resolution (r_R) in the slant range direction is

$$r_R = \frac{c\tau_p}{2} \tag{2.19}$$

Where: τ_p is the pulse duration.

Consequently the resolution (r_y) in the range direction (Figure 2.9) is given by

$$r_y = \frac{c\tau_p}{2\sin\theta} \tag{2.20}$$

Where θ is the angle of incidence.

The resolution in the azimuth direction is simply the arc length corresponding to a beam of width β_h, which is

$$r_a = \beta_h R = \frac{\beta_h h}{\cos\theta} \tag{2.21}$$

Where: h is the altitude of the radar.

From Equation 2.17 the angular resolution β_h is limited by

$$\beta_h = 1.22\frac{\lambda}{D} \tag{2.22}$$

so that

$$r_a \approx \frac{R\lambda}{D} \tag{2.23}$$

For an azimuth resolution of 10 m, using a satellite microwave sensor with an operating wavelength of 1 cm, operating at an altitude of 800 km, D is of the order of 100 m. It is not possible to fabricate a 100 m, space-borne, radar antenna with current technology. Thus to use radar from satellites it is necessary to "synthesise" a very large aperture (synthetic aperture radar, SAR). This is achieved by the satellite emitting a radar pulse and moving to new position to receive the return signal. This is

illustrated in Figure 2.11. The SAR is shown at three different positions along its flight path, A, B and C. At A, the forward edge of the radar beam is just intercepting the target, *T*. At B, the antenna is directly above the target, and at C the aft edge of the radar beam is just leaving the target. Since the synthetic aperture can only be built during the time that the target remains within the real antenna beam, the maximum possible synthetic aperture, L_p, is simply the azimuth resolution of the real aperture i.e.

$$L_p = \beta_{hr} R \tag{2.24}$$

Where: R is the range.

The subscript *r* is used to distinguish real aperture from synthetic aperture, *s*.

If the "real" aperture of the instrument is 10 m then, $\beta_{hr} \approx \frac{\lambda}{D} \approx \frac{1\text{ cm}}{10\text{ m}}$, and if the range $R_o \approx 800$ km, then the synthetic aperture $L_p \approx 800$ m.

In considering the beamwidth of a synthetic aperture, one must consider the phase shift in the radar signal between successive points along the synthetic aperture (Ulaby *et al.* 1982a). This phase shift occurs both in the downward and upward travelling radar pulses. It can be shown that the limiting angular resolution of a synthetic aperture, limited by this phase shift, is:

$$\beta_{hs} = \frac{\lambda}{2L} \tag{2.25}$$

Where: β_{hs} is the horizontal angular width of the radar synthetic beam.
L is the antenna length.

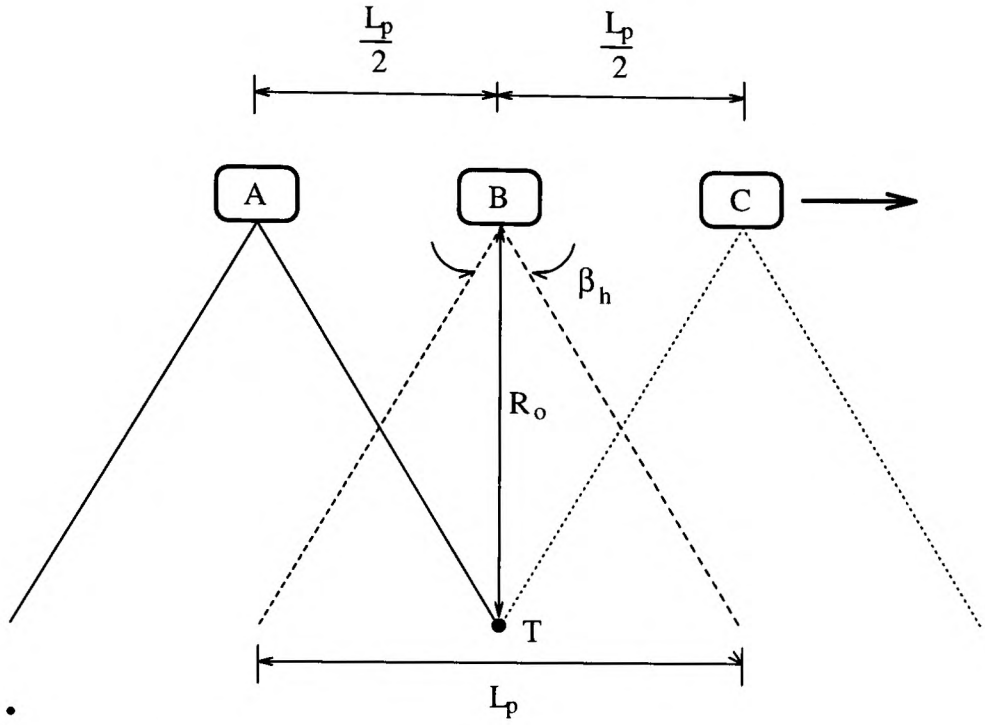


Figure 2.11: Radar beams from 3 points along a synthetic aperture A, B and C, illuminating a target, T at range R_o

The azimuthal resolution is distance along the Earth's surface intercepted by a beam of width $\beta_{hs}R$, which is given by:

$$\left[\text{using: } L=\beta_{hr}R \text{ and } \beta_{hr}\approx\frac{\lambda}{D}\right]$$

$$\beta_{hs}R=\frac{\lambda R}{2L}=\frac{\lambda R}{2\beta_{hr}R}=\frac{D}{2} \tag{2.26}$$

(Note that the definition of β_{hr} implies one way travel, while that for β_{hs} implies two-way travel.)

The signals received by the SAR at different point along the synthetic aperture, for a given target, require phase corrections because the different path lengths between the radar and the target result in different phase shifts. If the synthetic aperture is sufficiently small that these phase shifts are less than

$\pi/4$ radians they can be neglected, and in this case the radar is called unfocused. It can be shown that the aperture for an unfocused SAR satisfies:

$$L < \sqrt{\lambda R} \tag{2.27}$$

and from Equation 2.26, that the maximum resolution of unfocused SAR is reduced to:

$$\frac{1}{2}\sqrt{\lambda R} \tag{2.28}$$

Since radars usually look sideways, the range, R, is different for different viewing angles, and therefore the phase corrections are different for each range at a given azimuth.

An alternative way to think of SAR is as a "Doppler Beam Sharpener" (Ulaby *et al.* 1982a). Consider a radar travelling in direction x at velocity u transmitting an amplitude modulated radar signal,

$$S_t(t) = p(t)\cos(\omega_o t) \tag{2.29}$$

Where: p is the amplitude of the signal

ω_0 is the angular frequency

The reflected signal will have the form,

$$S_r(t) = \alpha p\left(t - \frac{2R}{c}\right)\cos\left[\omega_o\left(t - \frac{2R}{c}\right) + \varphi\right] \tag{2.30}$$

Where: α is the amplitude modification of the surface

φ is a phase shift, which is a property of the reflection process.

The transmitted signal is usually a series of pulses of duration t, at a repetition rate f where $t << f$. For a given target the frequency of the reflected signal changes as the radar moves because of the Doppler effect. It can be shown that the Doppler frequency shift (f_D) is given by:

$$f_D = \pm\frac{2u(x - x_o)}{\lambda R_o} \tag{2.31}$$

Where: x_o is position of the target,

R_o is the range to the target.

The Doppler frequency shift is positive for approaching targets, negative for receding targets. (Note that f_D is linear in x when $(x - x_o)^2 << R_o^2$).

In order to detect a frequency, f_D in an instrument where the signal is pulsed, the sampling theorem dictates that it is necessary to sample the signal at a frequency higher than $2 f_D$. This can lead to ambiguity problems with space borne SARs. The problem is that the pulses must be far enough apart in time, and therefore space, to prevent range ambiguities but the pulse rate must be higher than $2 f_D$ to prevent aliasing.

The quantity of data generated by SARs is enormous and poses a serious difficulty for the widespread use of radar remote sensing. It is usual for SAR systems to send their arrays of two dimensional (range and azimuth) data to ground stations for processing. The data from SARs are usually transmitted in a stream, sequential in range for one azimuth scan, before moving on to the next azimuth scan. It is therefore necessary to store data until processing of the whole images can be undertaken. The data processing can be carried out by

1. summing all the measurements and adding the appropriate phase correction to achieve focusing (this requires order N^2 operations), or
2. using the Doppler approach, by carrying out Fourier transforms on the data (order $N\log_2 N$ operations).

Since the Fourier method requires fewer operations, it is usually preferred. One fast Fourier transform (FFT) is required for each range point in one synthetic aperture (4000 range point in the case of the SEASAT SAR, launched in June 1978; SEASAT operated at 1.4 GHz, had a swath width of 100 km offset 250 km from the sub-satellite point, and had a spatial resolution of 25 m). The calculation for each range point is independent and can therefore be carried out in parallel.

Speckle, or target fading, is a major problem for SARs. It occurs when the target that represents a single pixel in the SAR image is composed of several point scatters (discrete objects such as buildings that have a strong backscatter signature). The signal received by the radar is a summation of the radiation from each of these point scatters. This summation will vary from point to point, and also at a fixed point from time to time whenever there is relative motion between the scatters and the radar. This occurs because an interference pattern is established because of the phase difference that results from the different path lengths between each point scatter and the radar. The effect is to produce random fluctuations in the return signal from area-extensive targets represented by one pixel. This results in an image with bright and dark pixels. Speckle can be a very serious indeed if only one independent sample of a target is observed of a given image. Speckle is greatly reduced if the resolution of the instrument is relatively large compared with the size of the point scatters; one can then average over various pixels to obtain an estimate of the scattering coefficient for an area.

Fortunately in most cases each individual pixel in a SAR image is composed of a number of independent samples. As more samples are averaged for each pixel the amount of speckle decreases. It can be shown that the number of independent samples for a pixel (N_a) is given by:

$$N_a = \frac{2r_a}{D} \tag{2.32}$$

where: r_a is the azimuth resolution of the radar,
D is the real aperture of the radar.

Careful design of the instrument is required to ensure that all independent samples are actually averaged.

A wide range of parameters affect the backscatter coefficient for soil and vegetation surfaces, but some generalisations are possible. The most important characteristics of the radiation are its frequency, polarisation and incidence angle; whilst those of the target are its biomass, dielectric constant (which is related to its water content), the geometry of the vegetation, and the morphology (roughness) and dielectric constant of the soil. Consequently the backscatter depends upon the vegetation type, species composition and physiological age, or on the soil type, degree of cultivation and its water content. Designing a radar system to estimate a particular surface parameter, e.g. soil moisture, is largely a problem of optimisation. The frequency, incidence angle and receive/transmit (rt) polarisation need to be specified such that the backscatter coefficient σ^o is strongly sensitive to soil moisture and minimally sensitive to soil roughness variations and to vegetation cover. The optimum configuration for determination of soil moisture has been found to be frequency in the 4-5 GHz range and angle of incidence 7-22°. Polarisation is important only if tillage practices are in use, when HV polarisation should be used because it is minimally sensitive to row direction (Ulaby *et al.* 1982b).

The penetration of microwaves through a canopy is dependent upon the wavelength and polarisation, and provides potential for studying not only the canopy, but the underlying soil and vegetation too. The optical thickness of the vegetation layer increases with increasing frequency of radiation, thus L band is influenced by the whole canopy, whilst X band observations are governed by the top layers only (Figure 2.12). There is no optimum wavelength for general studies because the choice of wavelength is application and species dependent. Sensitivity to vegetation type is enhanced when the wavelength used is approximately the same size as the canopy components because of resonance.

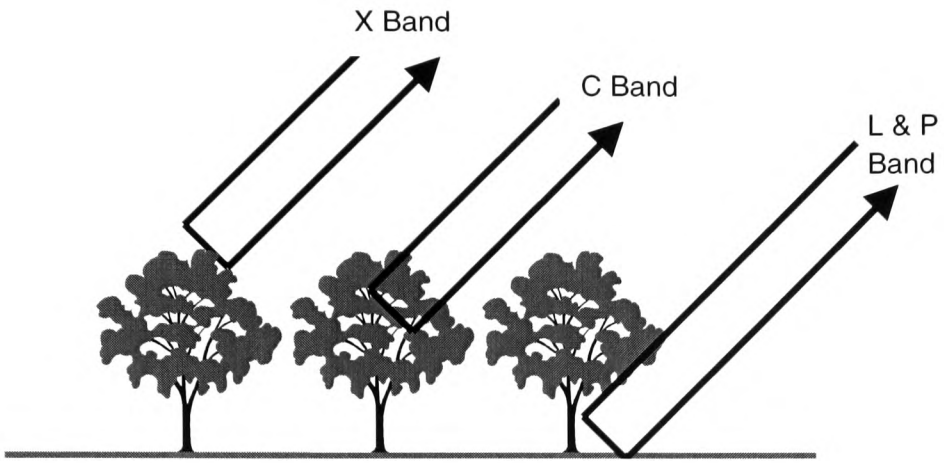


Figure 2.12 The primary interactions of X, C, L & P band microwaves with vegetation canopies.

Polarisation of the radar signal is very important. Horizontally transmitted, horizontally received radar (HH) couples weakly to vertical stalks, resulting in low attenuation and good penetration through the canopy. Vertically polarised, vertically received (VV) results in poor penetration. However, penetration depends strongly on frequency, therefore measurements using HH at L band can give information about underlying soil while VV will give data on canopy structure: discrimination of polarisation at X band becomes impossible. Most active microwave instruments have a fixed angle of incidence and, consequently, few studies of the effects of incidence angle have been made.

Crop species can greatly influence the backscatter signature, for example sugar beet (*Beta vulgaris*) gives a very strong return because it has a high water content, and because of its general geometry and large leaves (Ulaby *et al.* 1982). However, the reasons for the relationship between crop type and backscatter are not well understood. Multi-temporal studies greatly increase the classification rate for many crops since discrimination by radar tends to be site and time specific (Beaudoin, Le Toan and Gwyn 1990). Monitoring crop growth may be achieved by monitoring how backscatter changes with time, but there are species specific effects, e.g. for brassicas backscatter decreases with maturity, for wheat it increases.

Retrieval of LAI and Biomass

The sensitivity of NDVI to forest parameters is not high in canopies over eight years old when canopy coverage is typically achieved for many species. Moreover such techniques are limited because of the difficulty in retrieving parameters from mixed stands, and incompletely closed canopies and vegetation understory reflectance adds noise to the signal. Active microwaves techniques are more

promising for the retrieval of biomass because L and P bands are able to penetrate through leafy canopies, optical wavelengths are not. Recent studies have shown good results by combining X or C with L or P bands and using HV polarisation, the contrast between the different penetrative properties in these bands appears to improve detection efficiency (Ranson and Sun 1994). Ranson, Saatchi and Sun (1995) and Ranson *et al.* (1996) reported results showing that above-ground biomass could be estimated up to about 20 kg/m^2 ±1.5 kg/m^2 for the boreal forests of Canada. Although saturation at higher biomass density was not observed in these studies, poorer results above 20 kg/m^2 suggest degradation in accuracy above this density.

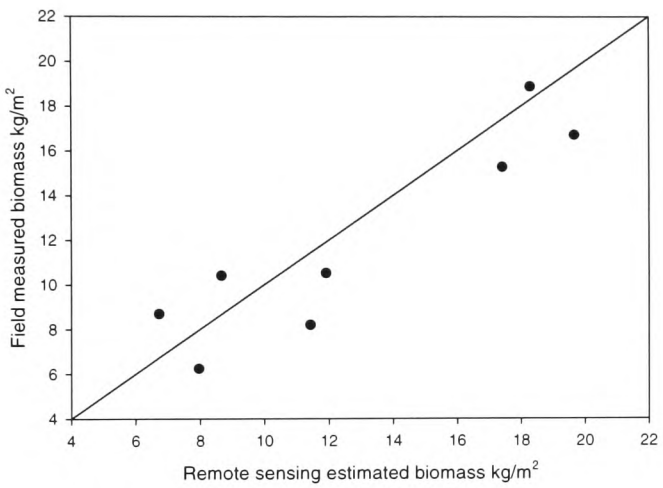


Figure 2.13 Scatter plot of field measured biomass against biomass estimated from L and C band SAR backscatter at HV polarisation. Redrawn from Ranson, Saatchi and Sun (1995)

Forest biomass frequently exceeds 15 kg/m^2 in areas such as tropical forests (40 kg/m^2) and the temperate rain forest of the Pacific Northwest (80 kg/m^2, Waring and Franklin 1979). Leaf area index is closely related to biomass, and consequently SAR estimation of LAI closely mirrors estimation of biomass

Retrieval of Soil Moisture

The radar backscatter signal is only sensitive to the moisture content of the upper soil layer (L band penetrates soil to a depth of about 10 cm). Sensitivity to soil moisture is severely attenuated by the effects of soil surface roughness (for example because of ploughing) and by the presence of a vegetation overstory (Beaudoin, le Toan and Gwyn 1990). Once the biomass of the vegetation

overstory exceeds about 0.1 kg/m^2, the ability of radar to retrieve information about the soil decreases rapidly. The amount of water in the leaves of a closed forest canopy is so large that it makes any direct assessment of soil moisture impossible.

Conclusion

Despite the complexities of SAR systems, the difficulties of interpreting the backscatter signal and the problems incurred processing the very large quantity of raw data which SARs produce, there appears to be scope for optimism regarding active microwave remote sensing. Good correlations have been found between biomass, LAI and radar backscatter. The development of backscatter models such as that of Sun *et al.* (1991) will improve our understanding of backscattering from vegetation canopies and lead to better sensor design. If invertable backscatter models can be developed, they will no doubt further improve the accuracy of parameter estimation using radar techniques. The power of computers is continually increasing and becoming more affordable. In the future parallel processing techniques may improve the efficiency of SAR processing dramatically. SAR sensors also have a significant advantage over optical sensors, in that they can operate even in cloudy conditions, thus enabling frequent, regular soundings, and enabling seasonal changes to be distinguished. With continuing development, particularly with multi-frequency SARs, the high spatial and temporal resolution offered by radars and their ability to penetrate clouds may soon enable determination of several useful surface parameters including biomass, surface roughness and soil surface water content. At present, the scarcity of datasets makes it difficult to evaluate the use of radar remote sensing for stand scale vegetation modelling.

2.3. Using Remotely Sensed Data in Models

The considerable problems associated with the derivation of parameters suitable for vegetation modelling from remote sensing become larger as the spatial scale of the model decreases from global to stand scale. Until recently, little thought was paid to the needs of environmental and ecological modellers when designing remote sensing instruments, with the result that the number, and range, of spectral channels, the spatial resolution of the sensor, and the orbit ephemeris of the satellite, have led to datasets which are far from ideal.

In the past, attempts to couple remotely sensed data with SVATs have concentrated on regional or global scale models. At these scales, errors in parameter retrieval from remote sensing are offset by the spatial heterogeneity of the landscape. Moreover, regional scale datasets are far more readily available than high-resolution data. Studies have mainly used the *forcing* approach where biophysical

variables are derived from satellite data either empirically, or through model inversion and then used as input to the vegetation model (Running 1988). In a classic example of the genre, Running *et al.* (1989) describe the regional scale inversion of LAI for a coniferous forest in Montana, USA. A second strategy for coupling remotely sensed data can be described as *calibration/validation.* In this approach, the canopy reflectance is derived from the vegetation model, which is then compared with the satellite signal. This comparison can be used to improve the performance of the model over time, thus achieving repeated re-calibration (an example of this technique was given by Nemani and Running 1989).

The parameterisation of vegetation models with remotely sensed data requires the accurate and frequent determination of several parameters, such as APAR, LAI, biomass, evapotranspiration, plant and soil water content, and phenological information such as timing of bud burst and senescence. Measurements must be insensitive to weather conditions, offer very high spatial resolution and be available daily, at least throughout the growing season. No one remote sensing technique by itself meets these requirements, however, the synergistic use of a combination of different techniques might enable remotely sensed data to be used, constructively, in models, by providing some information about at least some of the required parameters. Vegetation index datasets suffer from signal saturation and interference from changes in background reflectance but are widely available. BRDF datasets potentially offer more information about the vegetation canopy, but require complex models of radiation scattering, and datasets are scarce. However, BRDF modelling has been receiving increasing attention in recent years and the availability of datasets is likely to improve in the near future. Radar remote sensing offers many benefits compared to the more traditional optical techniques, but datasets are not generally available, and their use is still the subject of research. New techniques are being developed such as interferometric SAR and stereo-optical remote sensing, which offer further possibilities for parameter determination in the future. It is clear, therefore, that a new form of vegetation model is required that can take advantage of whatever remotely sensed data are available, whenever they are available, and make maximum use of any synergy between different techniques.

Chapter 3

Bidirectional Reflectance Distribution Function Models

3.1. INTRODUCTION

Optical wavelength remote sensing (RS) instruments have been used to study the Earth's surface for a number of years, and although active microwave remote sensing techniques are receiving increased attention, optical techniques remain by far the most common. Early RS instruments were single-look sensors that viewed their target at a single zenith angle, frequently at nadir. Unfortunately, complex three dimensional structures like vegetation canopies do not lend themselves to this type of observation as lower canopy layers quickly become obscured by layers above leading to the saturation of vegetation indices (Chapter 2). Moreover, the reflectance of light from vegetation canopies is highly anisotropic, consequently the signal received by a single look will vary significantly with the three dimensional geometry of the sensor-target-sun system. Recently multi-look sensors have been developed that observe the target from a number of different angles within a short space of time. The reflectance values measured by these instruments can be used to reconstruct the bidirectional reflectance distribution function (BRDF) of the target. This function contains a great deal more information about the surface than is available from nadir reflectance measurements. Unfortunately, in order to relate BRDF values to the canopy properties that are of interest to SVAT modellers, complex models of the interaction of electromagnetic radiation with vegetated surfaces are required. This chapter briefly reviews five BRDF models used in this study.

The spectral reflectance of vegetation canopies is the product of multiple scattering of direct solar radiation and radiation scattered by plant elements and the soil surface. The reflected radiance from a plant-soil system is directionally dependent, and the reflectance viewed by an observer varies with solar and viewing position. This bidirectional reflectance is a feature of almost all natural surfaces (Kimes and Sellers 1985). The use of remote sensing techniques for studying vegetation relies on the ability to relate reflectance measurements to vegetation properties. Canopy reflectance models play an essential rôle in achieving this objective by providing the logical connection between the biophysical properties of the canopy and the reflected radiation.

The solar radiation that reaches the Earth's surface is composed of direct solar beam and diffuse radiation that results from propagation through the atmosphere. The radiation that is reflected by the surface must propagate back through the atmosphere before reaching the sensor. For ground-based measurements the small volume of atmosphere between the sensor and the surface has little effect on the upwelling radiation. However, for aircraft and satellite-based sensors, atmospheric effects are significant in the optical wavebands and corrections for atmospheric effects are necessary. Atmospheric correction is a complex subject in itself, and was not addressed in this study. Consequently, it was assumed that all reflectance data have previously been corrected for atmospheric and other effects.

There are two aspects to the relationship between canopy parameters and reflected radiation measured by a sensor. The first involves an algorithm that will yield the set of measured reflectances given a set of system parameters. This is known as the forward modelling problem. The second aspect involves defining an algorithm that, given a set of measured reflectances, will generate the set of parameters that characterizes the canopy. This is known as the inverse modelling problem. Usually the solution to the forward modelling problem is a prerequisite for the solution of the inverse problem

3.2. CANOPY ARCHITECTURAL FACTORS

The two important architectural parameters of a canopy in terms of their influence on the BRDF are the leaf area index (LAI) and the leaf angle distribution (LAD). In the visible region leaves absorb most of the light incident upon them. Consequently, as the LAI of the canopy increases up to a certain value (usually about 2 to 3) most of the incident radiation is absorbed by the leaves and further increases in LAI have little effect on the canopy reflectance. In contrast, leaf absorption in the near infrared is minimal, so increasing LAI results in increasing radiation scattering resulting in increasing canopy reflectance until saturation occurs at an LAI of around 6 to 8. The LAD of a canopy is an import factor in determining the anisotropic nature of canopy reflectance, notably the hot-spot effect. The hot-spot is the peak in the canopy directional reflectance that occurs in the backscatter direction.

Figure 3.1 shows how the hot-spot effect arises, thorough a simple example of two flat leaves. Leaf 2 is nearly perpendicular to the direct sunlight and is therefore well illuminated. In contrast leaf 1 is nearly parallel to the sunlight and consequently poorly illuminated. An observer at B will see the dimly lit leaf 1 well, but cannot see very much of the brightly lit leaf 2. An observer at A with the sun behind, will see the bright leaves well, but cannot see very much of the dimly lit leaves. Thus to the observer at A the scene will appear very bright, whereas to the observer at B the scene will appear

darker. Therefore, when the sun is directly behind the observer (or sensor) a larger portion of the brightly lit foliage will be visible and shadows within the canopy, or on the soil surface, will be hidden. This is the hot-spot effect. Canopies are composed of many leaves with a range of different leaf inclination angles, therefore the magnitude of the hot-spot effect will depend on the LAD and the sun-vegetation-sensor geometry.

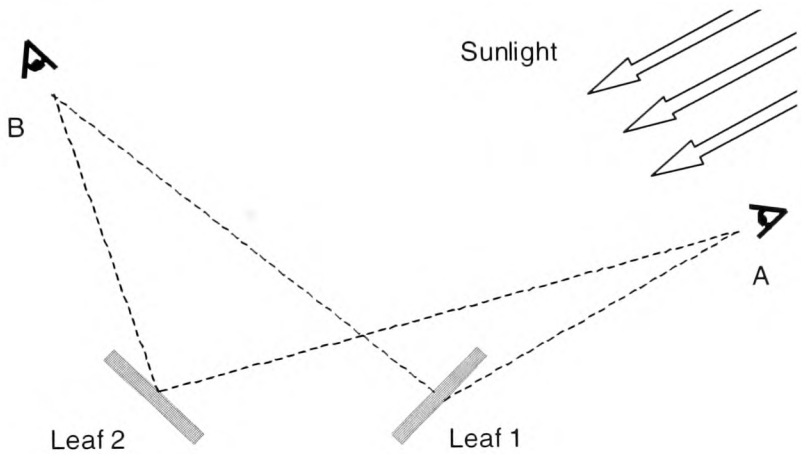


Figure 3.1 An illustration of how canopy geometry contributes to the bidirectional reflectance distribution function and the hot-spot. (After Norman, Welles and Walter 1985)

The recent years a great deal of work has been undertaken to understand and model the hot-spot. The hot-spot is considered important because it contains information about leaf size, orientation and overlap. However, in many cases the angular size of the canopy hot-spot is quite small, and consequently measurements of this effect from satellite sensors are difficult to obtain and infrequent.

3.3. MODELLING APPROACHES

A large number of models of the bidirectional reflectance of light from vegetation canopies have been developed in recent years. Whilst most of these models are based on the same fundamental physics, they vary greatly in the simplifications they employ. Models can be broken down into four general classes: two-stream approaches, radiative transfer and geometrical-optical models, although hybrid models have also been developed.

Two stream models

These models are based on average transmittance theory, in which the transmission of radiation as it travels through the canopy is related to the distribution and optical properties of the vegetation elements. The canopy is usually treated like a turbid medium where the vegetation elements are treated as small absorbing and scattering particles, with given optical properties and distributed randomly in horizontal layers and oriented in given directions. The canopy is treated as a horizontally uniform plane-parallel layer and the radiation field is assumed to depend only on the z co-ordinate. The earliest two-stream model was proposed by Suits (1972). In this model, the leaf elements are assumed to be oriented either vertically or horizontally. In keeping with the vast majority of models, leaves are assumed to have random orientation in azimuth. Leaf and soil optical properties are assumed to be Lambertian (i.e. perfectly diffuse). The SAIL model (Verhoef 1984) extends the Suits model to allow a discrete distribution of leaf orientation.

Radiative transfer model

This class of models was originally developed to predict the scattering properties of the atmosphere (Chandrasekhar 1950; Ishimaru 1987a,b) but in recent years radiative transfer models have been applied to vegetation canopies, e.g. Myneni (1990). These models have proved very accurate when properly calibrated, however, they are extremely computationally intensive and difficult to parameterize. Moreover, the complexities of radiative transfer modelling mean that these models are usually only one-dimensional, which implies horizontal homogeneity within the canopy, making these models poorly suited to modelling many canopies.

At the heart of radiative transfer theory is a differential equation (expressing the rate of change of radiation flux density in a given direction with position) that must be solved to calculate the reflectance of the canopy. This solution involves two steps, (i) the specification of the scattering properties of the vegetation elements in terms of the *phase functions* (the probability that radiance incident in a given direction will be scattered into the solid angle about a second direction), and (ii) the solution of the radiative transfer equation for specific boundary conditions. These steps are usually extremely difficult, but if the canopy is assumed to be a plane-parallel, infinitely extended medium, then the radiative transfer equations can be solved.

Geometric-optical models

In geometrical-optical models, the vegetation canopy is represented as a collection of discrete objects (e.g. tree crowns) on a plane. The reflectance is calculated as a function of the pattern of sunlit and shaded plants and soil visible from a given viewpoint. Simple geometrical models use geometrical optics to calculate the sunlit and shadowed regions of the scene and then use empirical relationships to determine the radiation regime within the canopy (e.g. Li and Strahler 1985). More sophisticated

models use simplified radiative transfer equations to determine the distribution of radiation within the canopy; consequently these models could be classed as hybrids.

Monte Carlo Models

Propagation of radiation in heterogeneous environments can be simulated as a random sequence of collisions between photons and obstacles in the form of vegetation elements. Monte Carlo methods can be used to estimate the statistical characteristics of this process (Ross and Marshak 1988). Each random event is simulated by the generation of a sequence of pseudo-random numbers according to the distribution density defined by the geometrical and optical properties of the canopy. That is to say the selection of random numbers is used to determine if a given photon will hit a vegetation element. If a vegetation elements is hit, the direction of the scattered radiation is again determined by a Monte Carlo procedure. Thus, the interception and scattering of radiation is followed numerically on a photon-by-photon basis and this make such models very accurate but extremely computationally intensive. Although the long computation time of Monte Carlo models precludes their use for many applications, they do provide a base-line for comparison with simpler canopy reflectance models. There are a number of difficulties associated with the field measurement of canopy reflectance data, including rapidly changing nature of the atmosphere, and problems encountered with the accurate positioning and pointing of the sensor. Consequently, high quality field measurements are difficult to obtain; in this context Monte Carlo prediction provides a valuable tool for validation of other canopy reflectance models

3.4. SAIL

The SAIL (scattering by arbitrarily inclined leaves) model (Verhoef 1984) is an extension of the earlier Suits model (Suits 1972a,b) to allow for any distribution of leaf inclination angle, and as such, is based on a two-stream Kubelka-Munk (KM) approximation to the radiative transfer equation in a turbid medium (Ishimaru 1987a). SAIL has been widely used by researchers for modelling canopy BRDF signatures because of its simplicity and its reasonably realistic representation of anisotropic reflectance. However, the assumption of a horizontally homogeneous canopy means that SAIL is best suited to modelling relatively uniform canopies such as crops, as opposed to forests. The canopy system is described by the following four differential equations giving the rate of change of radiation flux with vertical position x within the canopy:

$$\frac{dE_+}{dx} = -aE_+ + \sigma E_- + s'E_s \tag{3.1}$$

$$\frac{dE_-}{dx} = -\sigma E_+ + aE_- - sE_s \tag{3.2}$$

$$\frac{dE_s}{dx} = kE_s \tag{3.3}$$

$$\frac{dE_0}{dx} = wE_s + vE_- + uE_+ - KE_0 \tag{3.4}$$

Where: E_s is the downward direct solar flux

E_+ is the upward diffuse flux

E_- is the downward diffuse flux

E_0 is the internal radiance in the direction of observation.

k is the extinction coefficient for E_s

K is the extinction coefficient for internal radiance

a is the attenuation coefficient for diffuse flux

$\sigma, \sigma', s, s', u, v, w$ are various scattering coefficients corresponding to the different combinations of incident and scattered radiation flux, as shown in Table 3.1.

The calculation of the scattering coefficients is given in detail by Verhoef (1984) and solutions to equations 3.1, 3.2, 3.3 and 3.4 are given in Appendix A.

Flux	E_s	E_-	E_+
E_0	$w(\theta_l)$	$v(\theta_l)$	$u(\theta_l)$
E_-	$s(\theta_l)$	$\sigma'(\theta_l)$	$\sigma(\theta_l)$
E_+	$s'(\theta_l)$	$\sigma(\theta_l)$	$\sigma'(\theta_l)$

Table 3.1 The SAIL model scattering coefficients for the possible combinations of incident and scattered flux. For example, $s(\theta_l)$ is the coefficient for the scattering of incident solar beam flux, E_s, into downward diffuse flux, E_-.

Two versions of the SAIL model were developed in this study. The difference occurred in the way the leaf inclination angle distribution (LAD) was represented and consequently the parameter that formed the input to the model. The two versions of SAIL were otherwise identical. The two LAD representations used were the elliptical distribution of Campbell (1986) and the beta distribution of Goel and Strebel (1984). These two representations are discussed further in Chapter 7. The full list of SAIL input parameters is given in Table 3.2.

Parameter	Description	Parameter	Description
L	Leaf area index	ρ_s	Soil reflectance
ρ	Leaf reflectance	θ_s	Solar zenith angle
τ	Leaf transmittance	θ_o	Observation zenith angle
f_{sky}	The ratio of diffuse to total incident radiation	ψ	The relative azimuth angle between the sun and the observer
$\bar{\theta}$ or μ,ν	Mean leaf inclination angle, or the two parameters of the beta leaf angle distribution function.		

Table 3.2 The nine input parameters for the SAIL model. The leaf inclination angle may be represented, either by the elliptical leaf angle distribution (Campbell 1986) using a single parameter, the mean inclination angle, $\bar{\theta}$, or by a two parameter (μ,ν) beta distribution described by Goel and Strebel (1984). See also Equations 7.3 and 7.4.

3.5. GOEL AND GRIER ROW MODEL

Unfortunately, the assumption of a horizontally homogeneous canopy used in SAIL causes problems even for some types of crop canopy. This is because many agricultural crops tend to be planted in rows. The regular pattern of strips of bare soil, in between strips of crop canopy, has a significant effect on the BRDF. The Goel and Grier row model (ROW) (Goel and Grier 1986a,b) attempts to allow for these effects. ROW is a two-stream model based on the Suits row model (Suits 1983) but has two additional features. One of these features is the addition of a leaf inclination angle distribution and associated equations derived from the SAIL model. The second new feature is that the canopy is composed of elliptical sub-canopies (Figure 3.) arranged in rows of the form:

$$f(\delta)=\frac{1}{2}\left\{1-\left[\frac{\delta}{JP}\right]^2\right\}^{\frac{1}{2}} \qquad 0\le\delta\le JP \tag{3.5}$$

Where: δ is the across-row displacement

P is the row period

J is the fraction of the distance along P covered by half the plant canopy

In the early stages of growth J is small, and the canopy is open Figure 3.(a), but as the canopy grows J increases until the sub-canopies first touch each other ($J = 0.5$) and then overlap (Figure 3.(b)). Thus the surface is treated as a series of subcanopies separated by strips of bare soil, when $J < P/2$. Each subcanopy is treated as a turbid medium, and the SAIL model is used to describe the radiation regime

within these subcanopies. Row effects are introduced into the four differential equations for canopy flux of the SAIL model (Equations 3.1, 3.2, 3.3 and 3.4) by multiplying the scattering and attenuation coefficients by a row modulation factor $M(\delta)$, defined by Equation 3.6. In addition to J and P the ROW model introduces a third additional parameter to the nine parameters of the SAIL model, namely ψ_{ROAZ}, which represents the azimuthal direction of the canopy rows.

$J \leq 0.5$

$$M(\delta) = \begin{cases} A(\delta) & 0 \leq \delta \leq JP \\ A(JP) & JP \leq \delta \leq (1-J)P \\ 2A(JP) - A(P-\delta) & (1-J)P \leq \delta \leq P \end{cases} \tag{3.6}$$

$J \geq 0.5$

$$M(\delta) = \begin{cases} A(\delta) & 0 \leq \delta \leq \dfrac{P}{2} \\ 2A(JP) - A(P-\delta) & \dfrac{P}{2} \leq \delta \leq P \end{cases} \tag{3.7}$$

Where:

$$A(\delta) = \int_0^{\delta} f(\delta)\mathrm{d}\delta = \left(\frac{JP\pi}{4}\right) + \left(\frac{\delta}{2}\right)\left\{1 - \left(\frac{\delta}{JP}\right)^2\right\}^{\frac{1}{2}} - \left(\frac{JP}{2}\right)\cos^{-1}\left(\frac{\delta}{JP}\right) \tag{3.8}$$

and $f(\delta)$ is the height of the canopy at a distance δ from the centre of the canopy

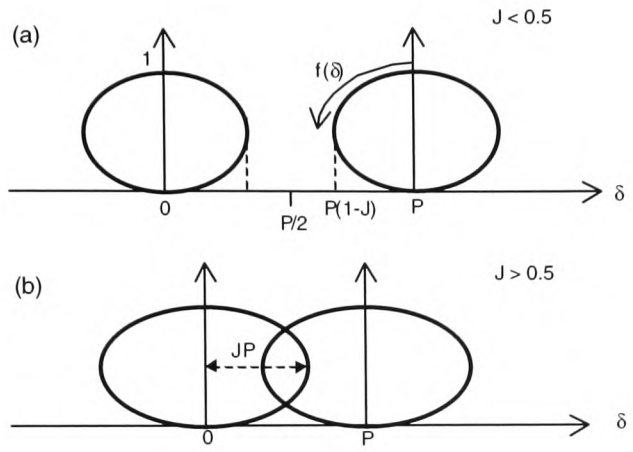


Figure 3.2 The ROW model canopy profile for two different growth stages. (a) An open canopy, where $J < P/2$. (b) A closed canopy, where $J > P/2$. (After Goel and Grier 1986a)

3.6. TRIM

The TRIM (three-dimensional radiation interaction model) model (Goel and Grier 1988) is an extension of the earlier ROW model and thus is a two-stream model with empirical corrections to account for the lack of horizontal homogeneity. In TRIM, the ground is divided into a rectangular grid with one axis (the η-axis) at an azimuth angle R_{oaz} and the other axis (the δ-axis) perpendicular to it. Thus it is designed to allow the modelling of vegetation composed of sub-canopies arranged in a grid pattern.

The sub-canopies are arranged at intervals P and Q times the canopy height (which is taken to be 1) along the δ-axis and η-axis respectively. The fractional distance along the δ-axis and η-axis covered by half the plant canopy is denoted by J and D. By varying J and D various canopy growth stages can be simulated including a row-planted canopy structure. The row modulation factor $M(\delta,\eta)$ for the TRIM model is given by:

$$M(\delta,\eta)=\begin{cases} C(\delta,\eta) & 0\le\delta\le P, 0\le\eta\le Q \\ C(\delta(\mathrm{mod}\,P),\eta(\mathrm{mod}\,Q)) & \text{elsewhere} \end{cases} \tag{3.9}$$

Where: $C(\delta,\eta)$ are continuous functions describing the profile of the sub-canopy.

For a scene composed of sub-canopies planted in a rectangular lattice structure with each tree planted at ($P/2$, $Q/2$), C is given by:

$$C(\delta,\eta) = \begin{cases} \sqrt{\left[1-\left(\frac{\delta-P/2}{JP}\right)^2-\left(\frac{\eta-Q/2}{DQ}\right)^2\right]} & \left[\left(\frac{\delta-P/2}{JP}\right)^2-\left(\frac{\eta-Q/2}{DQ}\right)^2\right]\le 1 \\ 0 & \text{elsewhere} \end{cases} \qquad (3.10)$$

The four equations 3.1, 3.2, 3.3 and 3.4, multiplied by the row modulation function 3.9 are solved in an analogous manner to the ROW model, except that the attenuation of direct flux is much more complicated because it involves determining the fraction of the incident ray that is within a given sub-canopy. Diffuse fluxes are assumed to independent of δ and η despite the density modulation. A solution to the above equations is given in the appendix of Goel and Grier (1988) and will not be repeated here.

3.7. LI-STRAHLER

The Li-Strahler model was originally proposed by Li and Strahler (1985). Since then the basic model has been modified, extended and improved on a number of occasions (Li and Strahler 1985, 1986, 1988, 1992; Strahler and Jupp 1990). The original (1985) model is a geometric-optical model that treats a forest stand as cones that cast a shadow on a contrasting background. This model has a number of limitations, in particular the individual cone shaped crowns are opaque, and the sensor is assumed to be near-nadir looking. Using the mean and variance of per-pixel reflectance, the model was successfully inverted to determine size and spacing of conifers without resolving them individually, provided that a homogeneous multi-pixel area could be identified *a priori* so that the variance in brightness values within the stand may be found.

Li and Strahler (1986) extended the 1985 model to arbitrary illumination and view directions making the model suitable for BRDF studies, and included an empirical correction for mutual shadowing effects. A translucent crown was also added with a negative exponential attenuation factor.

The mutual shadowing of crowns is an important phenomenon at large solar zenith angles because illumination shadows will tend to fall on other crowns, rather than the background, and will preferentially shadow the lower portions of adjacent crowns. Moreover, these shadows will tend to be

obscured themselves since adjacent crowns preferentially obscure the lower portions of other crowns. Consequently, at either large solar or viewing zenith angles, the tops of crowns are more likely to be illuminated and visible than lower portions, making the scene appear brighter (Figure 3.3). The effect leads to the characteristic bowl shape of the BRDF, in which scene brightness increases at large zenith angles.

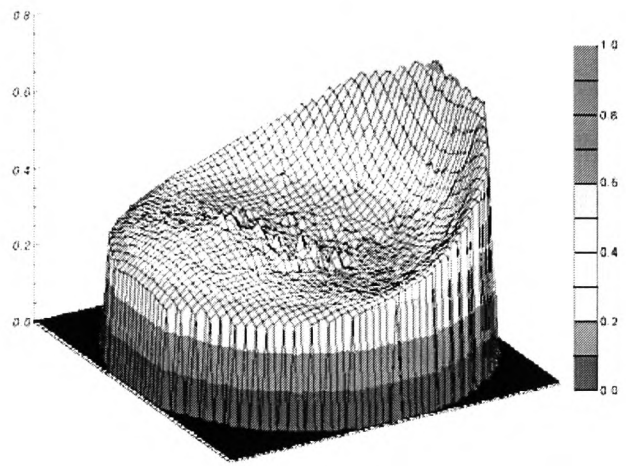


Figure 3.3 The BRDF for a rainforest in Brazil, exhibiting the characteristic "bowl" shape.

Li and Strahler (1992) improved upon the treatment of mutual shadowing in the 1986 model and also introduced a more general, ellipsoidal crown shape.

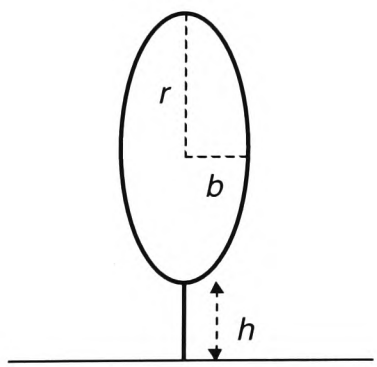


Figure 3.4 The elliptical crown geometry used in the later Li-Strahler model. The ellipsoid has a semi-major axis of length r and semi-minor axis of length b. The ellipsoid is positioned at a height h above the ground.

Trees are modelled as an ellipsoid with crown height $2r$, crown diameter $2b$ and h as the height from the ground to the bottom of the crown Figure 3.4. The signal received by the sensor is modelled as the area-weighted sum of contributions from four scene components: sunlit and shaded crown, and sunlit and shaded background. Therefore, the brightness of a given pixel in the image may be expressed as:

$$S = K_g G + K_c C + K_t T + K_z Z \tag{3.11}$$

Where: S is the brightness value of a pixel within the image

K_g is the areal proportion of sunlit background

K_c is the areal proportion of sunlit crown

K_t is the areal proportion of shadowed crown

K_z is the areal proportion of shadowed background

G is the spectral signature of sunlit background

C is the spectral signature of sunlit crown

T is the spectral signature of shadowed crown

Z is the spectral signature of shadowed background

The scene areal proportions are readily calculated in terms of r, b and h (Figure 3.4) from simple geometry, see Li and Strahler (1986, 1992). Using the Li-Strahler model the BRDF of an image pixel is modelled as the limit of its directional reflectance factor:

$$R(i,v) = \frac{\int\int_A R(s) f(i,s) f(v,s) I_i(s) I_v(s)}{A \cos\theta_i \theta_v} \mathrm{d}s \tag{3.12}$$

Where: $f(a,b) = \cos\theta_a \cos\theta_b + \sin\theta_a \sin\theta_b \cos\phi$

and $\mathrm{d}s$ is a small Lambertian surface element

A is the area of the pixel

$R(s)$ is the reflectance of ds

θ is the zenith angle of a direction

i, v, s represent the directions of illumination, viewing and the normal to the surface element ds, respectively

$I_i(s), I_v(s)$ are indicator functions equal to 1 if ds is illuminated or viewed

respectively, or 0 otherwise.[1]

The integrals in equation 3.12 are over the angles of illumination and viewing. The BRDF of a surface is usually defined for an infinite extent, however, in practice the sensor receives flux from only a limited area. In the above equation the sensor response is modelled as the spatial integration of the sensor footprint (A) over the x-y plane, hence the term A in the integral. Solutions to equation 3.12 can be formulated in terms of the scene spectral signatures and areal extents. The reader is referred to Abuelgasim and Strahler (1994) for an excellent summary of these formulations.

3.8. BÉGUÉ

The Bégué model attempts to model the radiatiave transfer in discontinuous clumped canopies in a realistic yet relative simple way. The canopy is simulated by a number of optically porous cylindrical objects spaced at regular intervals on a hexagonal grid. The model consists of two sub-models describing the macro-structure and micro-structure of the canopy. These sub-models adopt geometrical-optical and two-stream approaches respectively.

Macro-structure

The macro-structure model adopts a geometric representation of the spatial distribution of sub-canopies. Each cylinder has a height H and a radius R. The directional interception efficiency for a canopy of opaque cylinders has been shown to be equal to the ratio between the shadow cast by one cylinder per unit ground area (Brown and Pandolfo 1969), and can be expressed as

$$\varepsilon_i(\theta_s,\phi_s)=\frac{T+S-O}{A} \tag{3.1}$$

Where: T is the vertical projection of a cylinder and is given by: $T=\pi R^2$

S is the shadow area cast by a cylinder and is given by $S=2RH\tan\theta_s$

O is the area to be subtracted from S to allow for the shadows falling onto other cylinders.

θ_s,ϕ_s are the solar zenith and azimuth angles respectively

[1] In the case of translucent crowns, the indicator functions $I_i(s), I_v(s)$ are modified to be a function of the canopy transmittance and the path length through the crown to the surface element ds, see Li and Strahler (1986).

Bégué adapted the Brown and Pandolfo equations for the case of porous cylinders. The new directional interception efficiency is given by:

$$\varepsilon_i(\theta_s,\phi_s)=\frac{(T+S-O)(1-\rho)}{A}+\frac{\sum O_j p^{n_j-1}(1-\rho)}{A} \tag{3.2}$$

where the term $(1-\rho)$ represents the interception coefficient of the cylinder and the term on the right hand side allows for the second and subsequent interception of transmitted radiation by other cylinders. The term O_j is the fraction of the shadow area contributed by cylinder j and is a function of the cylinder spacing e. An algorithm to compute O_j is given by Bégué (1991).

Microstructure

The radiation regime within each cylinder is modelled assuming an infinite, homogeneous canopy of randomly distributed leaves. The gap frequency or porosity for such a canopy can be written as:

$$p(\theta_s)=\exp[-d(\theta_s)G(\theta_s)L_v] \tag{3.3}$$

Where: $G(\theta_s)$ is the average projection of unit leaf area.

$d(\theta_s)=\frac{H}{\cos\theta_s}$ is the optical path though a vegetation layer of height H

$L_v=\frac{L}{H}$ is the leaf area density.

The mean optical path through a cylinder $\bar{d}$ may be calculated by integrating d over the cylinder volume (height and radius). Consider the integration first over the cylinder height: the optical path $d(\theta_s,H,r)$ through a layer of height H width $2r$ and infinitesimal depth, then:

$$d(\theta_s,H,r)=\frac{1}{(H+h_l)}\int_0^{H+h_l} d(h)dh=\frac{2rH}{H\sin\theta_s+2r\cos\theta_s} \tag{3.4}$$

Where: $d(h)$ is the optical path at height h within the cylinder and

$$h_t = \frac{2r}{\tan\theta_s}$$

The radius integration is made over γ (see Figure 3.2) by making the substitution $r = R\cos\gamma$. This leads to:

$$\bar{d}(\theta_s, H, \eta) = \frac{8H\eta}{\pi}\int_0^{\pi/4} \frac{\cos\gamma}{\sin\theta_s + 2\eta\cos\theta_s\cos\gamma} d\gamma \qquad (3.5)$$

Where $\eta = \frac{H}{R}$ is the shape coefficient of the cylinder.

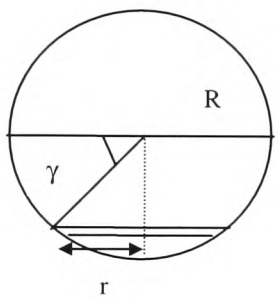


Figure 3.2 Geometry of integration over cylinder radius

Using the above equation 3.3 can be rewritten:

$$p(\theta_s) = \exp[-\bar{d}(\theta_s, \eta)G(\theta_s)HL_v] \qquad (3.6)$$

Where: $\bar{d}(\theta_s, \eta)G(\theta_s)$ is the extinction coefficient of the clump K_c

HL_v is the leaf area index of the cylinder

Spectral dependence is introduced by multiplying the extinction coefficient by the term $(1-\tau_F(\lambda))$, where $\tau_F(\lambda)$ is the transmittance of the leaves at wavelength λ. Thus, equation 3.6 becomes:

$$p(\theta_s) = \exp[-K_c(\theta_s, \eta)HL_v(1-\tau_F)] \qquad (3.7)$$

The incident diffuse radiation is assumed to be isotropic and its interception efficiency is calculated by integrating the direct interception efficiency over the upper hemisphere:

$$\varepsilon_{id} = \frac{2}{\pi}\int_0^{\pi/2} \varepsilon_i(\theta_s, \phi_s) d\theta_s \tag{3.8}$$

The canopy absorption efficiency can be written:

$$\varepsilon_a = \frac{\alpha_F}{(\alpha_F + \rho_F)}[\varepsilon_i + (1-\varepsilon_i)\rho_s \varepsilon_{id}] \tag{3.9}$$

Where: $(1-\varepsilon_i)$ is the incident radiation transmitted through the canopy

α_F, ρ_F are the leaf absorption and reflection coefficients respectively

ρ_s is the soil reflection coefficient

The canopy direction reflectance is calculated by assuming a horizontally homogeneous layer of transmittance $(1-\varepsilon_i) = \exp(-K'L)$, where K' is the canopy extinction coefficient, and L is the leaf area index of the canopy. The directional reflectance of the scene is composed of a canopy component and a soil component. The canopy component is given by integration of the elemental canopy reflectance from an infinitesimal layer of leaf area index *df* at a depth *f* in the canopy, by:

$$\rho_c'(\theta_s, \theta_v) = \frac{\rho_F}{\rho_F + \alpha_F}\int_0^F K' \exp(-K'f)\exp[-K'(\theta_v)f]df \tag{3.10}$$

Where: $\exp(-K'f)$ is the radiation incident on a layer

$\frac{\rho_F}{\rho_F + \alpha_F}$ is the fraction of the incident radiation reflected in the viewing direction θ_v

$K'df$ is a first order approximation to the probability of interception of radiation by a leaf

$\exp(-K'(\theta_v)f)$ is the fraction of the radiation reflected in the viewing direction that escapes though the canopy

Integrating equation 3.10 gives:

$$\rho_c'(\theta_s,\theta_v)=\frac{\rho_F K'}{(\alpha_F+\rho_F)[K'+K'(\theta_v)]}[\varepsilon_i+\varepsilon_i(\theta_v)-\varepsilon_i\varepsilon_i(\theta_v)] \qquad (3.11)$$

The soil component of the directional reflectance is given by:

$$\rho_s'(\theta_s,\theta_v)=\rho_s[1-\varepsilon_i(\theta_v)][1-\varepsilon_i][1-[\varepsilon_i+\varepsilon_i(\theta_v)]] \qquad (3.12)$$

Where: $1-\varepsilon_i(\theta_v)$ is the probability of viewing the soil

$1-\varepsilon_i$ is the probability that the soil is illuminated

$[1-[\varepsilon_i+\varepsilon_i(\theta_v)]]$ is the probability of viewing sunlit soil.

Finally the directional reflectance of the canopy is given by the sum of ρ_c' and ρ_s'.

3.9. CONCLUSIONS

The problem of modelling the anisotropic reflectance from vegetation canopies is a difficult one. The diverse range of different canopy structures, from relative homogeneous canopies such as those of some crops, to the discrete crowns of clumped leaves and branches that characterise many tree canopies, has lead to the result that BRDF models are frequently tailored to a particular canopy type. The exceptions to this rule are Monte Carlo models that are able to predict accurately the reflectance of different canopy types. Unfortunately, these models are too computationally intensive for widespread use other than to validate simpler models. Moreover, the inversion of Monte Carlo models is totally impractical in most cases. It is necessary, therefore, to compromise on model accuracy through in introduction of simplifications and assumptions, in order improve computation times.

The review of remote sensing techniques, and the applications of remote sensing to the derivation of parameters suitable for SVAT models described in Chapter 2, revealed that it is not possible to measure many of the parameters required by stand-scale SVATs using remote sensing techniques. Because of these difficulties, it was concluded that remote sensing techniques are best used to study changes over time that result from growth processes in vegetation canopies. Whilst optical remotely sensed data are abundant, traditional methods for using the data (e.g. vegetation indices) are prone to

signal saturation. Multi-angle techniques, such as BRDF measurement, offer several advantages over other optical techniques, but require the use of complex models of surface reflectance such as those described in this Chapter. Chapter 5 discusses two strategies for confronting SVAT models with remotely sensed BRDF data, and in Chapter 7, stochastic optimization techniques are applied to the inversion of the SAIL and ROW models. In Chapter 8 the development of a new SVAT model is described that incorporates both a BRDF submodel, and stochastic optimization subroutines.

Chapter 4

Leaf and Soil Reflectance Models

The leaf and soil reflectance, and leaf transmittance terms that are parameters of most BRDF models are dependent upon wavelength. Consequently in order to make use of multi-spectral remotely sensed data, models of the optical properties of leaves and soils are required that express the spectral reflectance in terms of wavelength independent parameters. These models may then be coupled to BRDF models in order to predict multi-spectral, bi-directional surface (canopy + soil) reflectance. However, leaf and soil optical properties appear as parameters in almost all BRDF models, and since these properties vary with wavelength, during any inversion of the BRDF model using multi-directional, multi-spectral data, separate values would have to be retrieved for each parameter for every wavelength. This would negate any benefit obtained from using multi-spectral data in the first place. In order to utilise directional, multi-spectral datasets, a BDRF model is required that contains wavelength independent parameters only.

4.1. THE PROSPECT MODEL

The PROSPECT model (Jacquemoud and Baret 1990) uses a radiative transfer approach to model the spectral reflectance of plant leaves. Using only three, wavelength independent, input parameters, PROSPECT can reproduce the reflectance and transmittance of healthy leaves to remarkably high degrees of accuracy. The absorption and scattering of electromagnetic radiation by plant leaves is a function of the leaf chemical composition and its physical structure. In the visible waveband, absorption results from transitions between orbital states of electrons in certain atoms - most notably chlorophyll, and some other pigments. A leaf's internal structure and the arrangement of tissues and cells results in refractive index discontinuities that cause internal scattering of radiation. In the infrared, the concentration and distribution of leaf water determine the absorptance. Although the internal structure of a leaf determines the optical properties throughout the whole spectrum, the effect is masked unless the absorptance is low (Gausman and Allen 1973).

When radiation is incident upon a leaf one fraction is reflected by the disc-like cells of the epidermis. The remaining radiation passes through the densely packed palisade parenchyma cells, which are arranged perpendicular to the leaf surface, into the spongy mesophyll. Mesophyll cells are much less densely packed and are interspersed with air spaces. The differences in refractive index between the hydrated cell walls and the intercellular air spaces cause the incident radiation to be scattered. Some of the scattered radiation escapes through the lower epidermis (transmitted radiation), the remainder

diffuses upward and exits through the upper epidermis, forming an additional component of the reflected radiation.

Several approaches have been adopted for modelling the optical properties of leaves including ray tracing, continuous approaches that use radiative transfer theory (e.g. Kubelka-Munk), and discrete approaches in which the leaf is modelled as a stack of opaque plates. In general, ray tracing approaches, as described by Allen *et al* (1973) and Brakke and Smith (1987), are too computationally intensive to be of widespread, practical use, and cannot be inverted to obtain leaf parameters from reflectance and transmittance measurements.

Allen and Richardson (1968) describe a model based on Kubelka-Munk theory that describes the leaf reflectance and transmittance with two input parameters - the leaf scattering and absorption coefficients. Baret *et al* (1988) showed how a simplified version of this model could be successfully applied to wheat leaves. Yamada and Fujimura (1988) improved the original Allen and Richardson model by considering leaves as being composed of four layers - two cuticles, a palisade parenchyma, and a spongy mesophyll.

The PROSPECT model is an improved version of the generalised "plate model" of Allen *et al* (1969, 1970). The plate model considers a leaf as a transparent plate with rough plane-parallel surfaces. The incident light beam is assumed to be perpendicular to the leaf surface. However, at the microscopic scale, leaf surfaces are not smooth (Grant 1987) and consequently incident radiation penetrates the leaf surface within a solid angle, Ω. The PROSPECT model modifies the plate model to take Ω into account.

In the original compact layer plate model leaf reflectance (ρ_a) and leaf transmittance (t_a) are given by

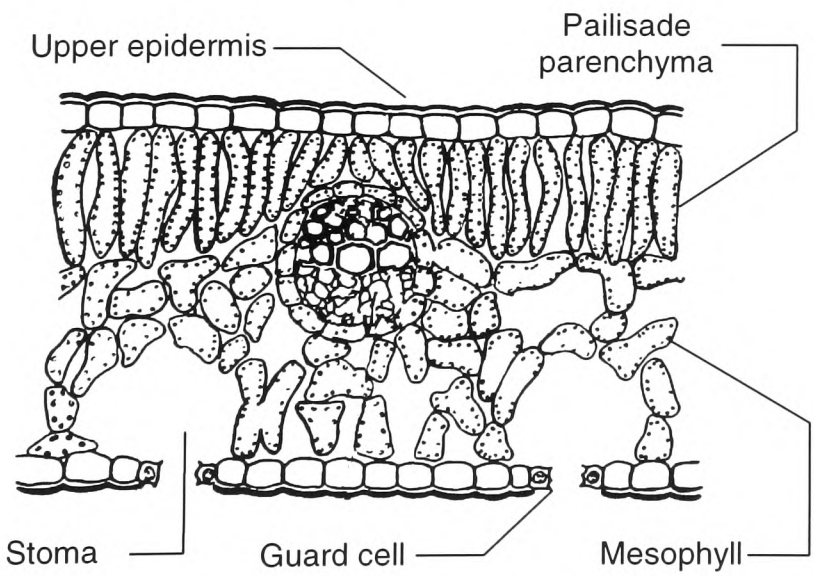


Figure 4.1 The internal structure of a typical amphistomatus leaf

(Allen and Richardson 1968):

$$\rho_a = \left[1 - t_{av}(\alpha,n)\right] + \frac{t_{av}(90,n)t_{av}(\alpha,n)\theta^2\left[n^2 - t_{av}(90,n)\right]}{n^4 - \theta^2\left[n^2 - t_{av}(90,n)\right]^2} \tag{4.1}$$

$$\tau_a = \frac{t_{av}(90,n)t_{av}(\alpha,n)\theta n^2}{n^4 - \theta^2\left[n^2 - t_{av}(90,n)\right]^2} \tag{4.2}$$

Where: α - is the maximum incidence angle, defining the solid angle Ω

n - is the refractive index

θ - is the transmission coefficient of the plate

The term $t_{av}(\alpha,n)$ is the transmissivity of a dielectric plane surface, averaged over all directions of incidence and over all polarisations. Its value can be expressed in terms of a very complicated equation which will not be shown here - see Allen (1973).

Using equations (4.1) and (4.2) we can write:

$$\rho_a = x\rho_{90} + y \tag{4.3}$$

$$t_a = xt_{90} \tag{4.4}$$

Where:

$$x = \frac{t_{av}(\alpha,n)}{t_{av}(90,n)} \tag{4.5}$$

$$y = x\left\{t_{av}(90,n) - 1\right\} + 1 - t_{av}(\alpha,n) \tag{4.6}$$

However, the assumption of a compact leaf does not hold for dicotyledonous or senescent leaves, and Allen *et al* (1970) extended the model into the so called generalised plate model by considering the leaf as being composed of a pile of N homogeneous layers, or plates, separated by $(N - 1)$ air spaces. The solution to the problem of light passing through a pile of plates has been known for over one hundred years and was first presented in a classic paper by Stokes (1862).

Stokes' solution assumes the light flux to be isotropic, whereas Jacquemoud and Baret (1990) modified this theory to include a non-diffuse beam incident on the leaf surface, i.e. the top plate of the pile. This is achieved by separating the top plate from the $(N - 1)$ others as follows. The top plate receives light at incident angle α, within a solid angle Ω. Let ρ_a and ι_a be the reflectance and transmittance of the top layer. Inside the leaf the light flux is assumed to be isotropic and the reflectance and transmittance of an internal leaf layer are given by ρ_{90} and ι_{90} respectively. The total reflectance $R_{N,\alpha}$ and transmittance $T_{N,\alpha}$ of a leaf composed of N layers, for light incident at an angle α to the leaf normal are given by:

$$R_{N,\alpha} = \rho_a + \frac{\iota_a \iota_{90} R_{N-1,90}}{1 - \rho_{90} R_{N-1,90}} \tag{4.7}$$

$$T_{N,\alpha} = \frac{\iota_a T_{N-1,90}}{1 - \rho_{90} R_{N-1,90}} \tag{4.8}$$

Using Equations (4.3) and (4.4) to eliminate ρ_a and ι_a we obtain:

$$R_{N,\alpha} = xR_{N,90} + y$$

$$T_{N,\alpha} = xT_{N,90}$$

Using the above equations to eliminate the terms for the reflectance and transmittance of the top layer, the inhomogeneous system can be transformed into a homogeneous system analogous to that used by Stokes (1862).

$$\frac{R_{N,90}}{b_{90}^N - b_{90}^{-N}} = \frac{T_{N,90}}{a_{90} - a_{90}^{-1}} = \frac{1}{a_{90} b_{90}^N - a_{90}^{-1} b_{90}^{-N}} \tag{4.9}$$

where

$$a_{90} = \frac{1 + \rho_{90}^2 - \iota_{90}^2 + \delta_{90}}{2\rho_{90}}$$

$$b_{90} = \frac{1 - \rho_{90}^2 + \iota_{90}^2 + \delta_{90}}{2\iota_{90}}$$

$$\delta_{90} = \sqrt{\left(\iota_{90}^2 - \rho_{90}^2 - 1\right)^2 - 4\rho_{90}^2}$$

The above series of equations requires four parameters, the incidence angle, α, the spectral refractive index, n, the transmission coefficient of the plate, θ, and the structure parameter N. Jacquemoud and

Baret (1990) did experiments on leaves of a number of plant species in order to determine α, and n . In the PROSPECT model the values of these two parameters are fixed to the values determined by experiment: α was fixed at 59 degrees, and n, which is a function of wavelength, is given by Figure 4.2.

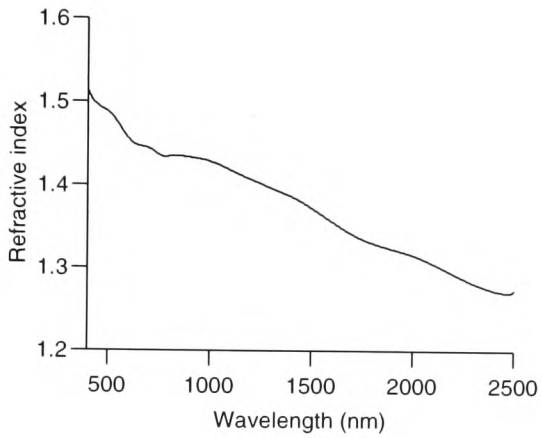


Figure 4.2 The relationship between refractive index (n) and wavelength used in the PROSPECT model, (Jacquemoud and Baret 1990).

To make the model useful, we need to relate the transmission coefficient to the leaf bio-chemical properties. From Allen *et al* (1969) the plate transmission coefficient θ is related to the spectral absorption coefficient $k(\lambda)$ by :

$$\theta - (1 - k(\lambda))e^{-k(\lambda)} - k(\lambda)^2 \int_{k(\lambda)}^{\infty} x^{-1} e^{-x} dx = 0 \qquad (4.11)$$

Suppose that the optical properties of the leaf are affected by several of its biochemical constituents, i. The spectral absorption coefficient $k(\lambda)$ can be written in terms of the specific absorption coefficients of the ith leaf bio-chemical components as follows:

$$k(\lambda) = \Sigma K_i(\lambda) C_i \qquad (4.12)$$

Where: λ is the wavelength

$K_i(\lambda)$ is the spectral specific absorption coefficient of component i

C_i is the content of leaf component i, per unit leaf area

It has been found from experiment that for albino and dry flat leaves where $C_i = 0$ (i.e. there is no pigment content), the absorptance is not zero (possibly because proteins, nucleic acids and other non-pigment leaf chemical constituents behave as light receptors (Maas and Dunlap 1989)).

Therefore a correction term $k_e(\lambda)$ is added to Equation (4.12) to account for this non-zero absorption effect, so that

$$k(\lambda) = \Sigma K_i(\lambda) C_i + k_e(\lambda) \tag{4.13}$$

Jacquemoud and Baret (1990), estimated $k_e(\lambda)$ from experimental data, and this is also incorporated into PROSPECT as a fixed parameter (Figure 4.3).

The PROSPECT model considers absorption by two components, absorption by water molecules within the leaf, and absorption by leaf pigments such as chlorophyll a and b and carotenoids. So Equation 4.13 can be rewritten as:

$$k(\lambda) = K_{ab}(\lambda) \cdot C_{ab} + K_w(\lambda) \cdot C_w + k_e(\lambda) \tag{4.14}$$

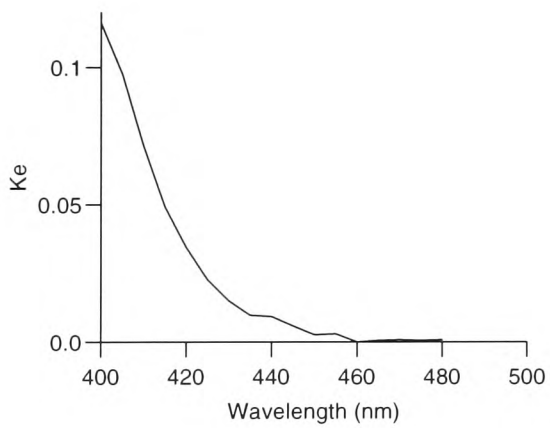


Figure 4.3 The relationship between the specific absorption coefficient of an albino leaf, k_e, and wavelength, as used in the PROSPECT model.

The values for K_w and K_{ab} are, again, found from experiment (Figures 4.4 and 4.5) and these values form a look-up-table that is used in the model. Therefore, the PROSPECT model has three wavelength independent input parameters, a leaf structure parameter, N, the chlorophyll a and b concentration C_{ab} and the leaf water content C_w.

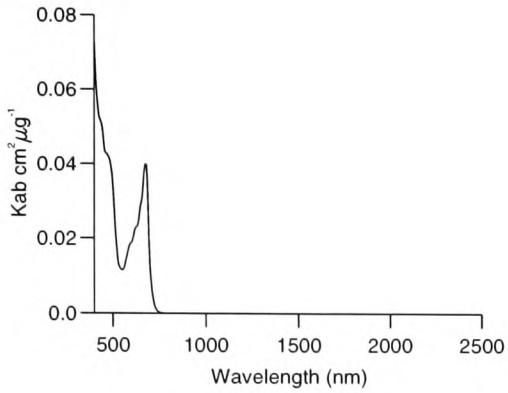


Figure 4.4 Specific absorption coefficient for chlorophyll aandb used in the PROSPECT model, K_{ab}, plotted against wavelength.

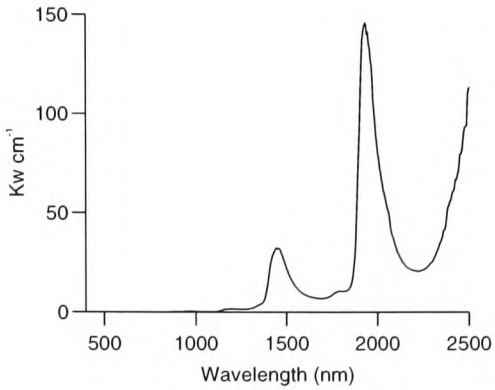


Figure 4.5 Specific absorption coefficient for leaf water used in the PROSPECT model, K_w, plotted against wavelength.

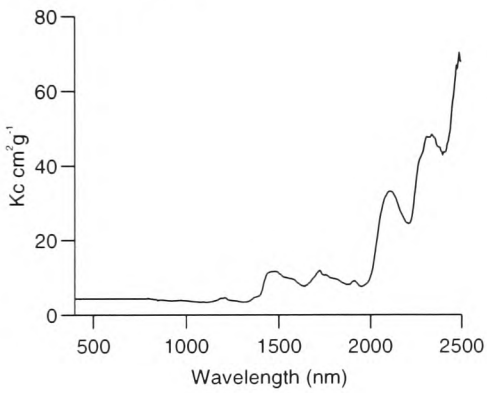


Figure 4.6 Specific absorption coefficient for cellulose and lignin used in the PROSPECT model, K_c, plotted against wavelength.

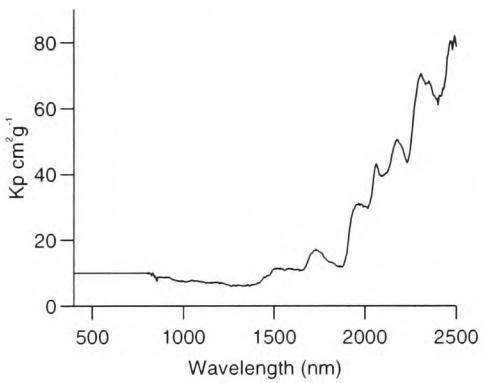


Figure 4.7 Specific absorption coefficient for leaf protein used in the PROSPECT model, K_p, plotted against wavelength.

In Figures 4.4 and Figure 4.5 it is clear that between 400 nm and 800 nm absorption is mainly caused by the presence of chlorophyll, whilst between 800 nm and 2500 nm, water is the dominant absorber. The presence of water in the leaf masks the much smaller absorption effects of cellulose, proteins and lignin (Peterson *et al* 1988), but, if the leaf is dry, absorption by these components may become significant. Consequently the PROSPECT model does not work well in these circumstances, and has recently been modified (Jacquemoud *et al.* in press) by introducing additional terms in Equation (4.13) for absorption by leaf proteins (K_p, C_p) and cellulose + lignin (K_c, C_c). The new model, PROSPECT Redux, is essentially identical to the original PROSPECT model, but now has five input parameters,: N, C_{ab}, C_w, C_c and C_p. Figure 4.6 and Figure 4.7, show the variation of the specific absorption coefficients for cellulose + lignin (K_c) and protein (K_p) with wavelength, determined from experiment

(Jacquemoud *et al.* 1998) and used in the PROSPECT Redux model. The use of a look-up table approach makes the PROSPECT model relatively fast. However, bearing in mind the fact is it is to be integrated into a BRDF model, a modular, optimised version of the PROSPECT model was created for use in this study to enable both the computation time to be minimised and the model to be linked to any BRDF model.

Figure 4.8, shows the sensitivity of predicted leaf reflectance (for a moist, healthy leaf) to variation of the parameters of the PROSPECT model. The reflectance is very sensitive to N, which is a measure of the structure of the leaf – principally its thickness. The average value of N for fresh leaves in typically around 1.5 for many species, but as the leaf dries out the number of air spaces within the leaf increases, effectively increasing N. As one would expect, the sensitivity of leaf reflectance to N is highest in the near-infrared region, where scattering of incident light is dominant, and minimal at the red-edge around 720 nm where absorptance is maximal. Sensitivity to chlorophyll (C_{ab}) is confined to the visible region of the spectrum as illustrated in Figure 4.8b, and reflectance is inversely proportional to chlorophyll content. Increasing water content reduces leaf reflectance above 1000 nm, with the most significant effect occurring in the water absorption bands near 1400 nm and 1900 nm. For fresh, moist leaves there is minimal sensitivity to C_c and C_p within the region from about 750 nm to 1300 nm and outwith the water absorption bands beyond 1600 nm. Figure 4.13 shows the same sensitivity plots for leaf transmittance. Equations 4.7 and 4.8 show that PROSPECT handles reflectance and transmittance in a near symmetrical way. This symmetry is illustrated in Figure 4.13b,c,d,e; these plots for transmittance are the same as those for leaf reflectance in Figure 4.8. However, Figure 4.13, shows that leaf transmittance decreases as N increases – the opposite behaviour to leaf reflectance. Figure 4.14 shows the sensitivity of modelled dry leaf reflectance and transmittance to variation of model parameters specifically connected to leaf moisture status, C_w, C_c and C_p. In Figure 4.14a and b, the leaf water content varies around a small mean value (this is a dry leaf), and therefore changes in optical properties are only noticeable near the water absorption bands. The model shows a higher sensitivity to the C_c and C_p parameters for dry leaves. An order of magnitude change in C_c produces a change in reflectance of 10% at around 2100 nm; but on the near-infrared reflectance plateau, the reflectance change is only 5%. The dynamic range for C_p is much smaller, and the maximum sensitivity is once again for wavelengths longer than 2000 nm, with a doubling of C_p concentration resulting in variations in reflectance of less than 1% at 1000 nm.

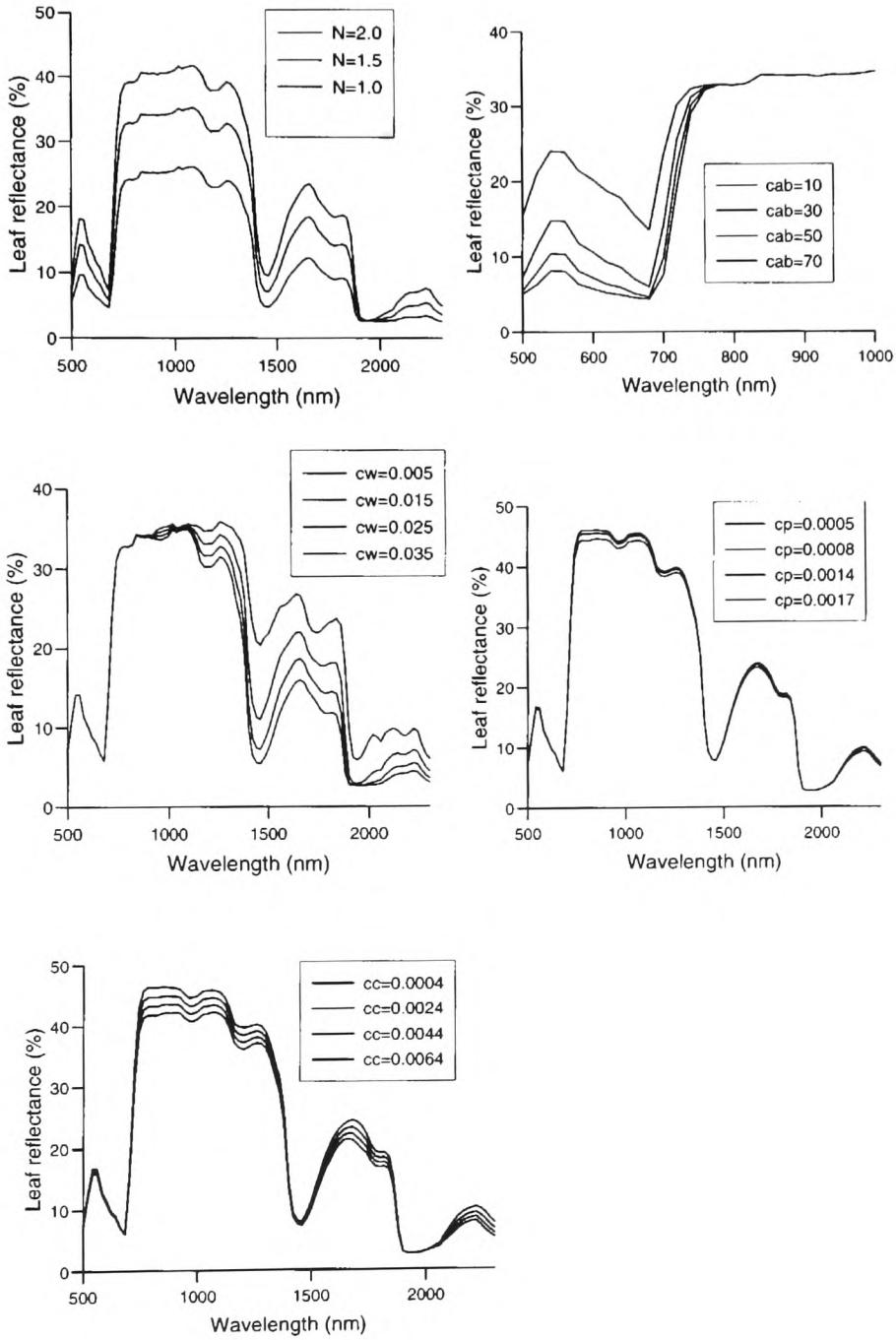


Figure 4.8 Sensitivity of PROSPECT modelled leaf reflectance to changes in the model input parameters varying around mean values of N=1.5, C_{ab} = 35 µg cm^{-2}; C_w = 0.020 cm, C_p = 0.0011 g cm^{-2}, C_c = 0.0034 g cm^{-2}. Range of values chosen to mirror the range found in typical leaves.

Inversion of PROSPECT would potentially allow retrieval of leaf bio-chemical information from measured leaf reflectance. The model sensitivity analysis indicates that N, and C_{ab} should be easily retrieved by inversion, however the insensitivity of the model to cp and cc for fresh leaves is so small that accurate estimation of these parameters is unlikely to be achievable. Sensitivity of dry-leaf reflectance to these parameters is somewhat higher, but possibly still not sufficient to allow retrieval of these parameters with confidence. Unfortunately, studies of vegetation reflectance using remote sensing instruments are usually at the canopy (or larger) scale and this introduces two problems. Firstly, a canopy reflectance model must be used to take into account the scattering effects of leaves and soil. Secondly, the difficulties in making atmospheric corrections and other factors mean that the spectral range of many instruments is confined to the visible and near-infra red part of the spectrum at wavelengths of less than 1200 nm. Thus, datasets frequently do not include the spectral regions where optical properties show highest sensitivity to C_w, C_c, and C_p.

Figures 4.9 and 4.10 illustrate the ability of PROSPECT to simulate accurately the reflectance and transmittance of healthy plant leaves. In each of these figures PROSPECT was inverted on measured data using the simulated annealing technique (see Chapter 6) to determine the best fit model parameters. The parameter set was then used as input in forward running the model to produce the curves. Figure 4.9 shows the inversion of PROSPECT on data for soybean (*Glycine max*) measured by Ranson *et al* (1982) and supplied as part of the NASA BRDF database (Ranson, Biehl and Bauer, 1985). The modelled data fit extremely well, with a slight under estimation of reflectance and transmittance in the near infrared plateau region, and a somewhat sharper change from low to higher reflectance and transmittance across the red edge than indicated by the measured data.

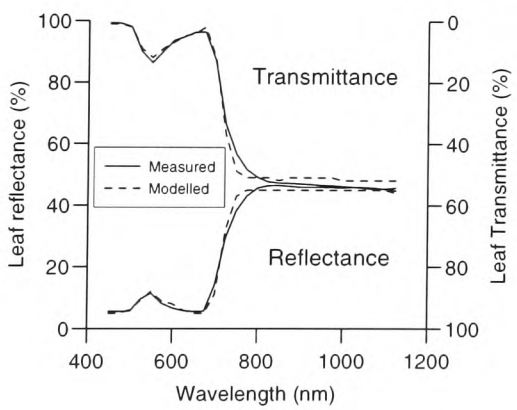


Figure 4.9 Comparison between reflectance and transmittance for soybean modelled by the PROSPECT model, and data measured by Ranson *et al.* (1982), NASA BRDF database.

Figures 4.10a and 4.10b show PROSPECT modelled reflectance and transmittance for aspen (*Populus tremuloides*) leaves measured by Walter-Shea *et al.* (1994) during the BOREAS project.

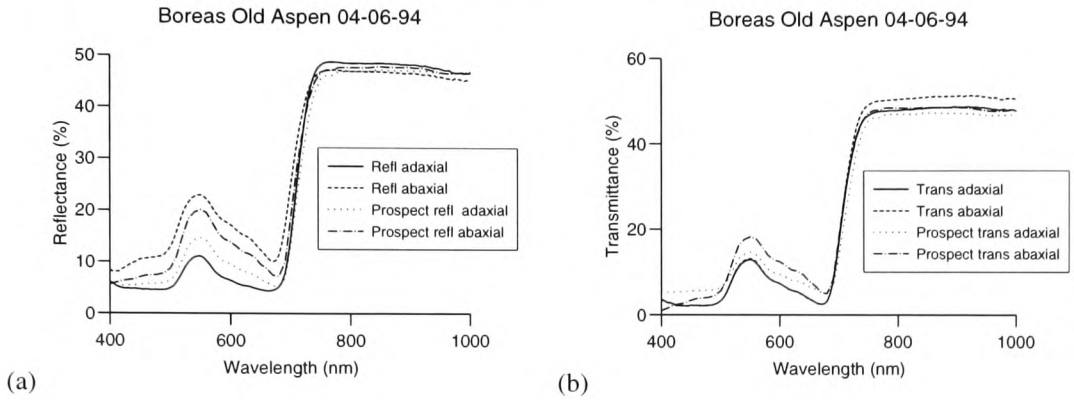


Figure 4.10 Comparison of PROSPECT modelled reflectance (a) and transmittance (b) for aspen leaves (adaxial and abaxial surfaces), with measured data (Walter-Shea *et al.* (1994), BOREAS project).

The PROSPECT modelled transmittance values show a very close fit to the measured data, with the largest discrepancies occurring between 400 nm and 500 nm, where the model tends to overestimate the transmittance slightly. The abaxial surface transmittance is also over estimated by the model in the near infrared plateau region. Errors in PROSPECT modelled leaf reflectance in the near infrared are very small, but exhibit a maximum error of around 4% near 550 nm. The transmittance of the abaxial and adaxial leaf surfaces are very similar, but there is a larger variation in the reflectance values. The PROSPECT model does not allow for different reflectance and transmittance values on the leaf abaxial and adaxial surfaces. In terms of canopy reflectance modelling, this is unlikely to pose a problem because this simplifying assumption is made in many BDRF models. Moreover, for many plant species, it is the adaxial leaf surfaces that are exposed to the most light, and at PPFD wavelengths most of the light incident on a leaf is absorbed. In the near infrared, scattered light may interact with the abaxial or adaxial surfaces of several leaves before emerging from the canopy, thus reducing the effect of difference in optical properties between the two leaf surfaces.

The PROSPECT model is less successful in modelling the optical properties of old, senescent or unhealthy leaves. Figures 4.11a and 4.11b show PROSPECT modelled leaf optical properties for old and senescent millet (*Pennisetum spp.*) leaves respectively. Measured data were collected by N. Hanan during the HAPEX-SAHEL project. Figure 4.11a shows that PROSPECT models the transmittance of old millet leaves well, but overestimates the transmittance by almost 100% near 550 nm. Reflectance is modelled less well, with PROSPECT underestimating leaf reflectance between 400 nm and 700 nm, indicating that the model is not able to represent the changes in reflectance that

occur through changes in physiology or biochemical constituents, as the leaf ages. This is reinforced by Figure 4.11b, which shows the modelled *vs* measured comparison for a senescent millet leaf. From this figure it is clear that PROSPECT is unable to represent the continuous increase in leaf reflectance from 400 nm to 1000nm and the absence of a red edge that characterise the observations. There is also a marked overestimation of leaf transmittance in the infrared.

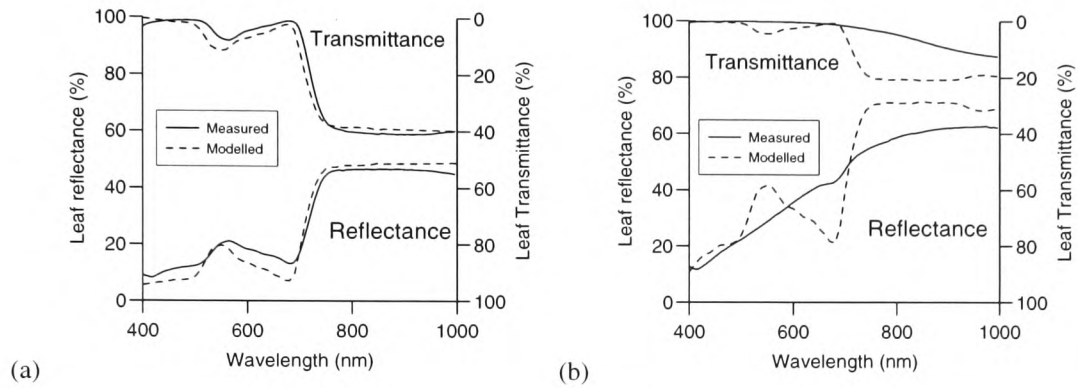


Figure 4.11 Prospect modelled optical properties for old (a) and senescent (b) millet leaves, compared with data measured by N. Hanan, HAPEX-Sahel project. (Data obtained form the HAPEX-Sahel database.)

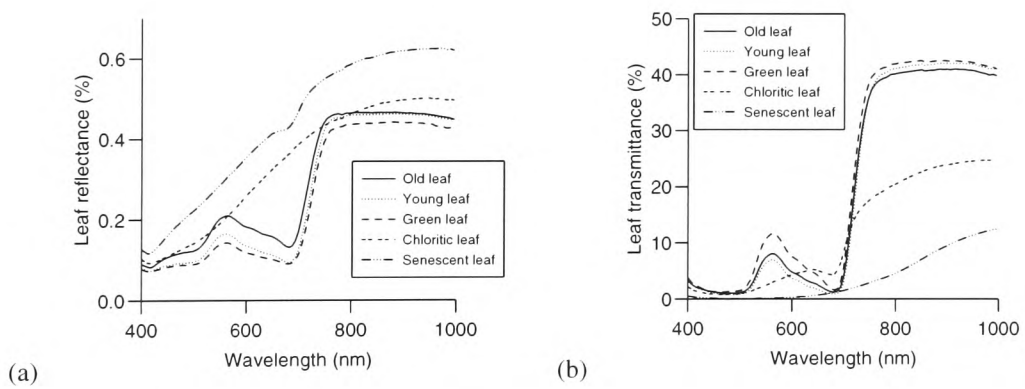


Figure 4.12 Measured reflectance (a), and transmittance (b), for millet leaves in various stages of health. Data collected by N. Hanan, HAPEX-SAHEL project.

Figures 4.12a and 4.12b compare the leaf reflectance and transmittance of millet leafs at various stages of growth and health. As the leaf ages, the reflectance and transmittance in visible wavelengths increases reflecting the physiological changes occurring and, in particular, the diminishing absorption of light by chlorophyll. During senescence, changes in the physical structure and water content of leaves, in addition to further changes in biochemical composition, result in radical changes to leaf

optical properties, characterised by a disappearance of the red edge and a significant increase in opacity.

Conclusion

The PROSPECT model, is a simple, fast and elegant mordel of leaf optical properties, capable of accurate representation of reflection and transmission spectra of the leaves of a wide variety of plant species. By comparing Figures 4.11 and 4.12, one can see that PROSPECT is able to represent the optical properties of healthy leaves with some accuracy, and excellent results are possible through inversion of the model on measured data. However, the changes which occur as the leaves near senescence, or suffer from serious stress leading to chlorosis, change spectral reflectance in such a way that inversion of PROSPECT on these data leads to poor model representations of the observed spectra.

4.1.1. Linking PROSPECT and SAIL

To enable the use of multi-spectral remotely sensed data with a BRDF model, the input parameters of the model must be wavelength independent. Unfortunately, models such as SAIL include leaf transmittance and reflectance as input parameters, although both these parameters are wavelength dependent. However, the PROSPECT model is able to compute leaf optical properties based on four wavelength independent input parameters. Although coupling the BRDF and PROSPECT models, and replacing the leaf reflectance and transmittance by the four parameters of the PROSPECT model increases the number of model parameters by two, and further increases the complexity of the model, it does enable the use of multi-spectral data, thereby increasing the available information content of the remotely sensed signal over that of multi-directional data alone.

In this study, the SAIL and PROSPECT models were coupled to form a new model, PROSAIL. Table 4.1 shows the input parameters of the coupled model. The parameters of the PROSAIL model are wavelength independent with the exception of soil reflectance and the fraction of diffuse radiation, f_{sky}. Thus PROSAIL is a BRDF model capable of modelling multi-spectral directional reflectance provided the spectral response of soil reflectance and diffuse radiation are known. Figures 4.15 and 4.16 show the sensitivity of PROSAIL modelled canopy reflectance to the parameters controlling leaf optics (for canopies of fresh and dry leaves, respectively in nadir viewing). The sensitivity of PROSAIL modelled canopy reflectance to changes in these parameters is generally less than the sensitivity of PROSPECT modelled leaf reflectance to the same parameters. This is because a canopy is composed of an ensemble of leaves with different positions and orientations. Light reflected or transmitted by one leaf may interact with other leaves in the canopy, being absorbed, reflected, transmitted or scattered by other leaves, before (possibly) escaping from the canopy. This multiple interaction

reduces the sensitivity of canopy reflectance to changes in the optical properties of the individual leaves.

Another parameter affecting the canopy reflectance is the reflectance of the soil background (Running 1991). Figure 4.17 shows the soil reflectance used in the simulations for Figures 4.15 and 4.16. The soil reflectance increases steadily through the visible and NIR regions from about 5% at 400 nm to 32% at 1700 nm. A water absorption peak is visible near 2000 nm. Near the red-edge at 760 nm, the soil and leaf reflectances are approximately the same size. In contrast to leaf reflectance, which is characterised by a relatively flat plateau region in the NIR, soil reflectance in NIR increases slightly just beyond the red edge reaching a maximum near 1200 nm.

The influence of soil reflectance on the canopy reflectance is also dependent on canopy structure, (in particular the leaf area index, leaf area density distribution and leaf angle distribution) viewing geometry and wavelength.

	Leaf area index	L	
	Leaf structure	N	PROSPECT model parameters
	Leaf chlorophyll content	C_{ab}	
	Leaf water content	C_w	
	Leaf protein content	C_p	
	Leaf cellulose and lignin content	C_c	
	Mean leaf angle	$\overline{\theta}$	
	Soil reflectance	$\rho_s(\lambda)$	Wavelength dependent parameters
	Diffuse radiation fraction	f_{sky}	
Sun-sensor geometry parameters	Solar zenith angle	θ_s	
	View zenith angle	θ_o	
	Relative azimuth angle	ψ	

Table 4-1 Input parameters of the PROSAIL model.

The fraction of incident radiation that is diffuse, f_{sky}, is invariant for solar zenith angle less than 60° (Monteith and Unsworth, 1990). However, it does vary strongly with the atmospheric conditions, and wavelength (Figure 4.12c). The direction of direct flux is characterized by the solar zenith and azimuth angles, whereas the diffuse flux is characterized by its spatial distribution – this is frequently approximated to be Lambertian[2]. The asymmetrical shape of the BDRF of vegetation canopies with

[2] See Monteith and Unsworth (1990) for a detailed treatment of this issue.

azimuth angle results from geometrical-optical effects caused by mutual shading of leaves in direct light (the hot-spot). Therefore, as the fraction of diffuse light increases, so does the azimuthal symmetry of the BRDF. Clevers and Verhoef (1991) illustrated that the influence of f_{sky} on SAIL simulated reflectances is only minor. Moreover, optical remote sensing measurements are best made under relatively cloud free conditions to both aid atmospheric correction and avoid cloud-pixel contamination. Consequently, the value of f_{sky} for most datasets lies in the range 1.0 to 3.0. Therefore, in this study, f_{sky} in the PROSAIL model, and variants, was replaced by a wavelength-independent parameter V_{is} related to the horizontal visibility. V_{is} was determined from a look-up table derived using the 5S model (Tanré *et al.* 1990).

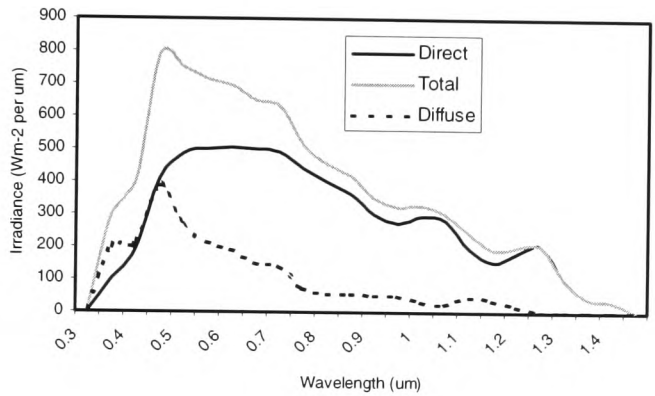


Figure 4.12c The spectral distribution of total, direct and diffuse solar radiation calculated for a model atmosphere by Tooming and Gulyaev (1967). Solar elevation = 30°, air mass number, m = 2, precipitable water 21mm. *(*After Monteith and Unsworth , *1990).*

4.2. Soil Reflectance Models

Soil reflectance can be a very significant component of the total surface reflectance measured by a remote sensing instrument. Its significance depends upon the vegetation structure and viewing geometry, as outlined above, and on the degree of canopy coverage, since remote sensing image pixels may contain reflectance contributions from several different surface components. In order to make use of multi-spectral remotely sensed data, a model of soil reflectance is required for which the parameters are wavelength independent. Unfortunately, such models are not common. Research has mainly concentrated on developing soil BRDF models, e.g. Pinty *et al.* (1989), Jacquemoud *et al.* (1992). For the sake of simplicity, many BRDF models assume a Lambertian soil background reflectance, which can, of course, be easily derived from soil BRDF models. The problem remains how to predict the spectral reflectance over a wide wavelength range for a number of different soil

types, from sand to peat, different soil textures, from smooth and fine to coarse and granular, and different soil moisture contents, from wet to dry.

4.2.1. The SOILSPECT model

One of the most successful soil reflectance models, SOILSPECT (Jacquemoud *et al.* 1992), is based on the work of Hapke (1981) and considers the soil surface as being composed of irregular and randomly oriented particles that are large compared with the wavelength of the illuminating light. The radiance I seen by a detector is considered as the sum of single scattering, I_s and multiple scattering, I_m terms:

$$I = I_s + I_m \tag{4.1}$$

and

$$I_s = \frac{J\omega}{4\pi}\frac{\cos(i)}{\cos(i)+\cos(e)}P(g) \tag{4.2}$$

$$I_m = \frac{J\omega}{4\pi}\frac{\cos(i)}{cs(i)+\cos(e)}\left[H(\cos(i))H(\cos(e))-1\right] \tag{4.3}$$

$$H(x) = \frac{1+2x}{1+2\sqrt{1-\omega x}} \tag{4.4}$$

Where: J is the intensity of illuminating light

i is the illumination zenith angle,

e is the view zenith angle

ω is the ratio of the scattered energy to the total energy either scatter or absorbed by a soil particle – called the *single scattering albedo.*

$P(g)$ is phase function, which describes the angular distribution of scattered light.

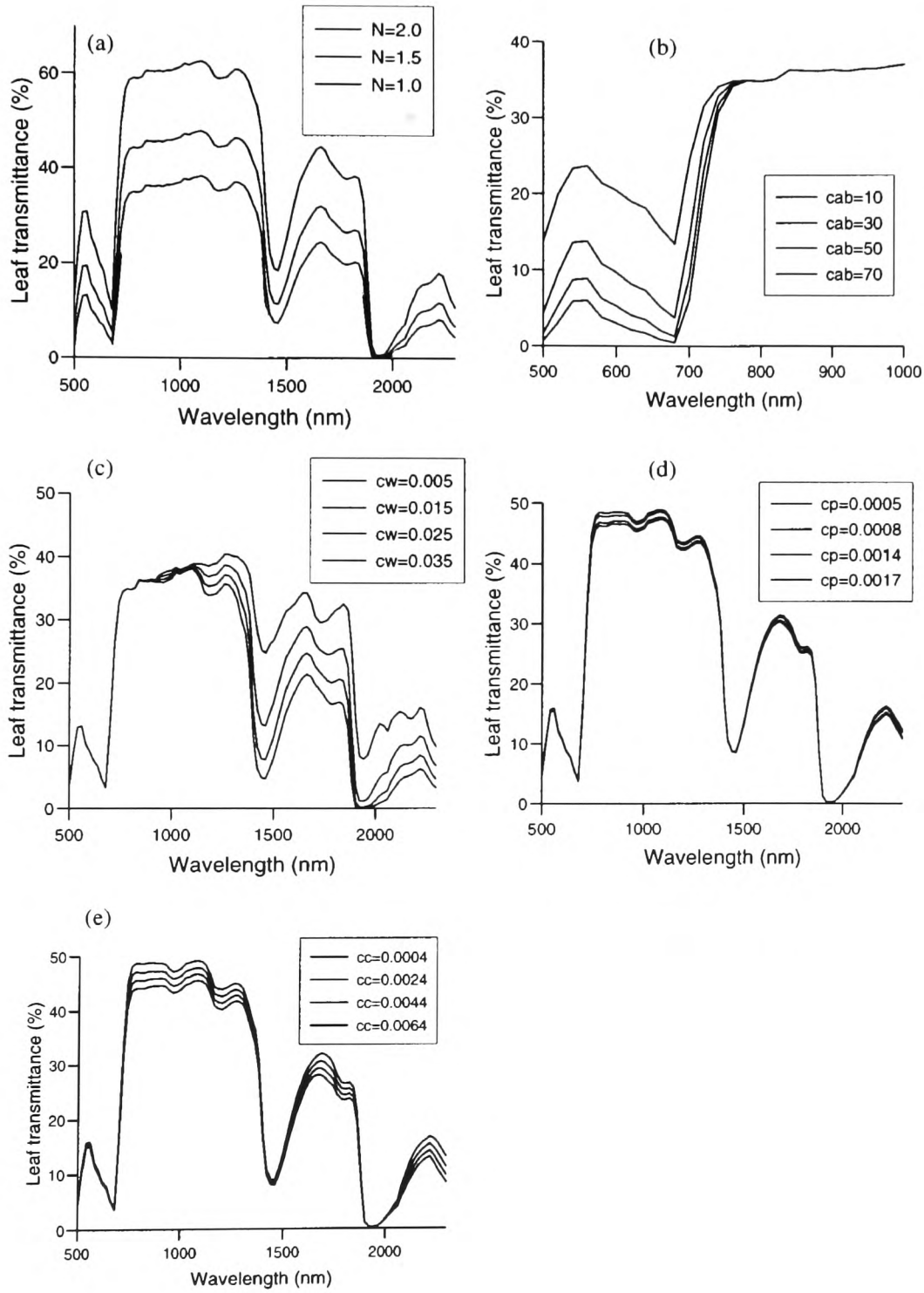


Figure 4.13 Sensitivity of PROSPECT modelled leaf transmittance to changes in the model input parameters. Mean value of parameters as in Figure 4.8.

In his classic work, Chandrasekar (1960) showed that (for a semi-infinte medium) the sensitivity of multiply scattered radiation to the particle phase function is much less than that for single scattered radiation. Therefore, the SOILSPECT model calculates I_s exactly for any phase function P and phase angle g, but assumes isotropic scattering for I_m ($P(g) = 1$). In early studies in this field (e.g. Hapke 1981), formulations of the phase function have corresponded to just one type of scattering – Lambertian, predominantly forward scattering or predominantly backscattering. The SOILSPECT model uses a phase function $P(g,g')$ approximated by Legendre polynomials to explain both forward and backward scattering, with the angle g' describing the angle between the forward scattering and the outgoing light directions, as follows:

$$P(g,g') = 1 + b\cos(g) + c\frac{3\cos^2(g)-1}{2} + b'\cos(g') + c'\frac{3\cos^2(g')-1}{2} \quad (4.5)$$

$$\cos(g) = \cos(i)\cos(e) + \sin(i)\sin(e)\cos(\phi) \quad (4.6)$$

$$\cos(g') = \cos(i)\cos(e) - \sin(i)\sin(e)\cos(\phi) \quad (4.7)$$

Where ϕ is the relative azimuth angle

b,b',c,c' are empirical parameters controlling the partitioning of reflection between forward, backward and mixed scattering.

Although the above equations work well for smooth soils, an additional term is added to explain the reflectance properties of rough soils, and to account for the hot spot effect (which occurs when $g = 0$). Following the technique of Hapke (1963), the SOILSPECT model includes a backscattering function, $B(g)$, which is a function of the phase angle, g, and a soil roughness index, h, as follows:

$$B(g) = \frac{1}{1 + \frac{1}{h}\tan\left(\frac{g}{2}\right)} \quad (4.8)$$

Thus, Equation 4.1 can be rewritten:

$$I = \frac{J\omega}{4\pi}\frac{\cos(i)}{\cos(i)+\cos(e)}\left\{\left[1 + B(g)\right]P(g) + H(\cos(i))H(\cos(e)) - 1\right\} \quad (4.9)$$

Equation 4.9 shows that, for a given phase angle g, increasing h leads to an increase in the backscattering function, corresponding to the increased bacscattering observed for rough soils. The SOILSPECT model, therefore, requires six input parameters the roughness index h, four parameters governing the directional reflectance, and the single scattering albedo ω.

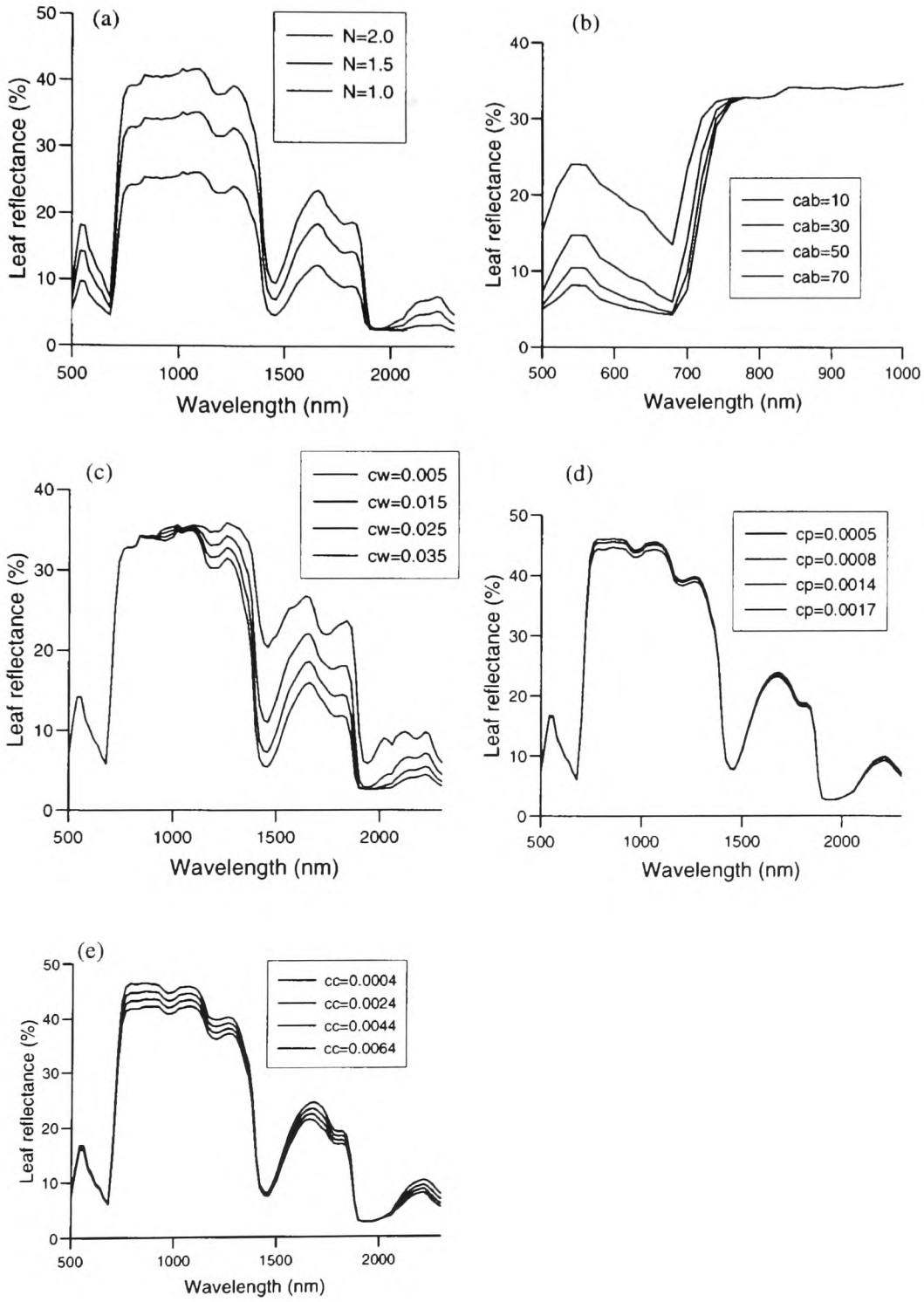


Figure 4.14 PROSPECT modelled leaf reflectance (a-c) and transmittance (d-e) for dry leaves Mean values for parameters: N= 2.2, C_{ab} = 30 μg cm^{-1}, C_w = 0.0005, C_p = 0.001 g cm $^{-1}$, C_c = 0.002 g cm^{-1}.

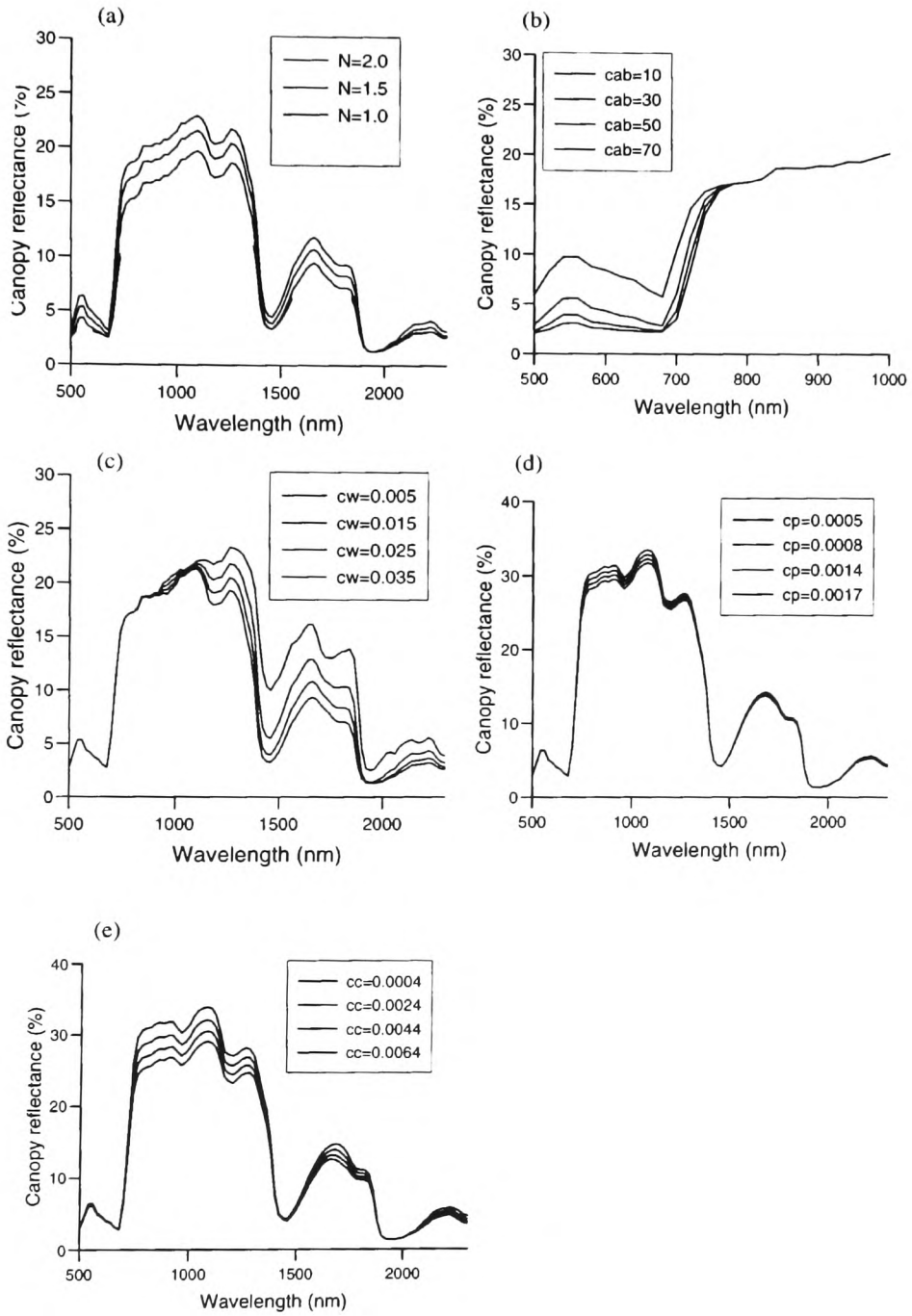


Figure 4.15 PROSAIL modelled reflectance for canopy of healthy moist leaves. Parameters as in Figure 4.8 with LAI =2.

To test the performance and invertability of the model, I inverted SOILSPECT on a measured dataset using simulated annealing. The model parameters derived from the inversion were then used for forward running of SOILSPECT in order to carry out a modelled/measured comparison. Jacquemoud *et al* (1992) and Pinty *et al* (1989) reported that the phase function parameters, *b, b', c, c'*, and the roughness index, *h*, are almost independent of wavelength, but the single scattering albedo, ω, is strongly dependent on wavelength. This leads to difficulties when inverting the model: for each wavelength in the dataset, a different value of ω has to be fitted in addition to the five other parameters. Therefore, sufficient directional data must be available, in addition to any spectral data, to provide sufficient measurements on which to base the inversion. For satellite sensors the number of directional measurements is probably restricted to no more than a handful of different zenith angles, the azimuth angle effectively being fixed by the satellite orbit ephemeris.

Figures 4.18 and 4.19 show a comparison between simulated and measured bidirectional reflectance for soil in HAPEX-Sahel. The measured data were obtained from the HAPEX-Sahel information system, and were measured by Alfredo Huete, *et al.* using a Spectron CE590 radiometer. Although measurements were ground based, the amount of directional information was limited. In general, measurements were taken at five different zenith angles in three different azimuths. The dataset used contained measurement for soils with two different soil moisture contents at wavelengths between 450 nm and 900 nm. Inversion was first attempted using the downhill simplex technique (Figure 4.18). This technique produced good results in several cases, resulting in good fits between modelled reflectances derived from the values of the parameters obtained in the inversion and the measured values. However, the inversion was found to be unstable, failing to converge in several cases. The scatter of points in Figure 4.18 also indicates either poor parameter estimation or that the model is not able to represent the bidirectional reflectance accurately. Examination of the inversion results in more detail revealed that, in many cases, although the model fit was poor, the final value of the merit function obtained during the inversion was small and of the same order as inversion runs where the resulting model fit was good. This type of behaviour indicates that the inversion was repeatedly becoming trapped in local minima: the results from the same inversion performed with simulated annealing (Figure 4.19) confirm this hypothesis. Figure 4.19 shows that inversion by simulated annealing produced excellent results, indicating that the model is invertable, even with this limited dataset.

The HAPEX dataset contained measurements made at different soil moisture contents, loosely defined as wet and dry. Results from this inversion confirm the findings of Jacquemoud *et al* (1992) that the parameters *b* and *c* are insensitive, whereas ω is strongly sensitive, to soil moisture. Parameters *b'* and *c'* show slight sensitivity to soil moisture and can take on both positive and negative values. This

variation might be explained by the low weight of these parameters in the phase function (Equation 4.5) and the resulting low sensitivity of the inversion of these parameters. The response of the model to changes in soil moisture is in line with the hypothesis that if the soil dries without significant changes in its mechanical properties then the phase functions parameters, representing the soil roughness and structure, remain constant. The single scattering albedo, ω, represents the intrinsic optical properties of the soil material, and hence is strongly sensitive to soil moisture, but is independent of the illumination and viewing angles. Figure 4.20 shows the effect of different soil moistures on ω for three different soil types, using values of ω obtained from simulated annealing inversion of SOILSPECT on data from Baumgardner *et al* (1985). The overall shape of the soil spectrum varies with soil type according to the nature of the soil composition. Clay, sand and peat exhibit an increase in albedo as the wavelength increases from the visible to infrared region. Peat which has a much larger organic matter content than the other two soil types, reflects very little in the visible region, whereas its reflectance increases greatly in the infrared region. Water absorption bands are clearly visible near 1450 nm and 1950 nm for all three soil types. As soils become drier, there is an increase in ω across the whole spectrum, corresponding to the widely observed increase in brightness of dry soils compared with wet soils.

Modifying SOILSPECT to Include the Effect of Soil Moisture: SOILSPECT+

In order to use SOILSPECT with multi-spectral soil reflectance data, the variation of ω throughout the wavelength range must be known, because ω is not wavelength independent. Consequently, in any inversion of the model it is desirable to have a wide range of directional data available. As indicated above, this may pose a practical problem for satellite remotely sensed data. A look up table could be constructed for ω based on a small range of soil types, but Figure 4.20 shows that ω is also a function of soil moisture, making the creation of such a table difficult. Modelling soil spectral reflectance as a function of water content is a difficult task. The commonly observed darkening of rough surfaces such as soil when they become wet was explained by Ångström (1925). He proposed that the diffuse reflection of light at a rough surface leads to total internal reflections at the liquid-air interface of the layer covering the surface. This reflection increases the probability of absorption of light by soil particles, expressed through the equation:

$$\rho_{wet} = \frac{\rho_{dry}}{n^2(1-\rho_{dry})+\rho_{dry}} \tag{4.10}$$

where: ρ_{wet} is the reflectance of a wet surface

ρ_{dry} is the reflectance of a dry surface

n is the refractive index of water.

Angstom's wavelength independent approach can be made wavelength dependent by making n a function of wavelength (Lekner and Dorf, 1988). Between 450 nm and 800 nm water absorption is

negligible and this simple equation works quite well. Beyond this wavelength range, particular water absorption bands exist. These features were modelled by Palmer and Williams (1974) using Lambert's law modified for two-way travel (Equation 4.11), and by assuming absorption occurs in an effective water layer within the soil.

$$R_{wet} = R_{dry} e^{\alpha l} \tag{4.11}$$

where: R_{wet} is the reflectance of wet soil

R_{dry} is the reflectance of dry soil

α is the spectral absorption coefficient of water

l is the active thickness of the water layer

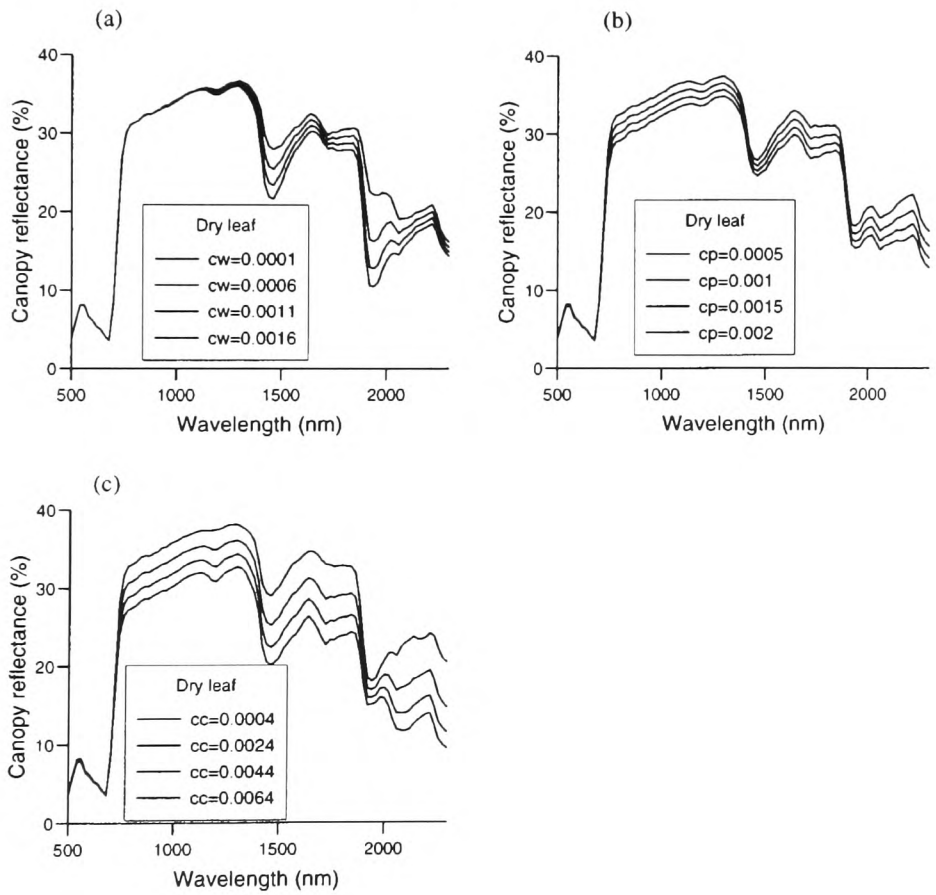


Figure 4.16 PROSAIL modelled canopy reflectance for dry leaves. Parameters as in Figure 4.14, with LAI = 2.

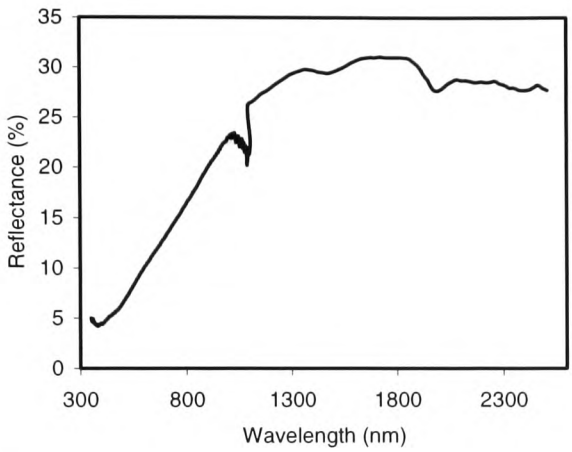


Figure 4.17 Soil reflectance used for the PROSAIL model simulations in Figures 4.15 and 4.16. Data obtained form the FIFE database (NASA) (Sellers *et al.* 1992).

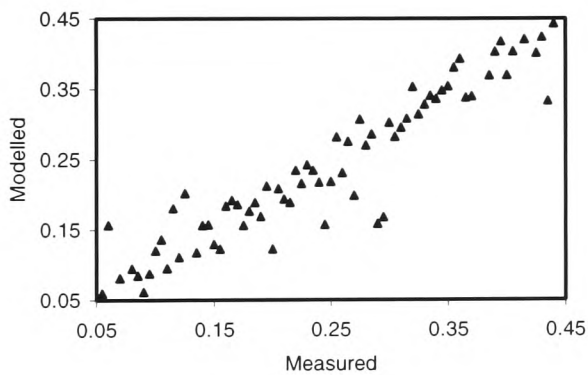


Figure 4.18 Comparison between modelled and measured bidirectional soil reflectance for HAPEX-Sahel. Downhill simplex inversion method.

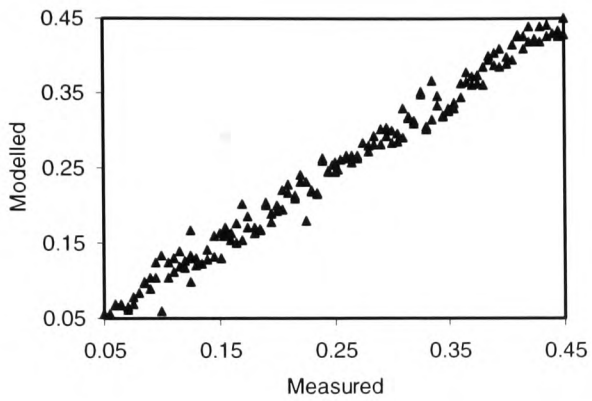


Figure 4.19 Comparison between modelled and measured bidirectional soil reflectance for HAPEX-Sahel - simulated annealing inversion method.

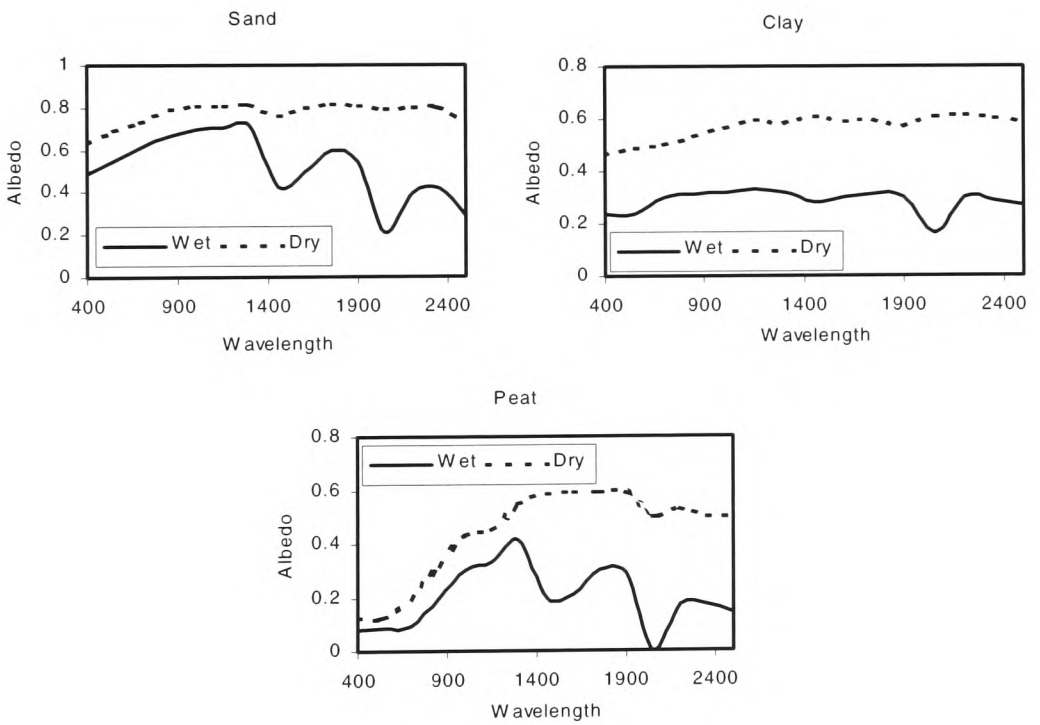


Figure 4.20 The single scattering albedo, ω, for wet and dry soils of three different types.

The water in moist soils is bound to the soil particles, and setting α to the value of the spectral absorption coefficient for pure water does not yield good results (Lekner and Dorf 1988). Lambert's law (Equation 4.11) assumes optically thin layers, and it can be expected that any deviation from this

assumption increases with increasing water layer thickness (i.e. soil water content) and therefore increasing absorption. This hypothesis is confirmed by the observation that the modelled reflectance for wet soil using the absorption coefficients for pure water deviates significantly from the measured values. The problem may be overcome by determining the absorption coefficients, α, from inversion of the soil reflectance model given by Equations 4.10 and 4.11 for a controlled dataset consisting of soils with known water content and dry reflectance values. Using data from Mauser (1996) the model was inverted on measured soil reflectance data using the standard simulated annealing technique (see Chapter 7)

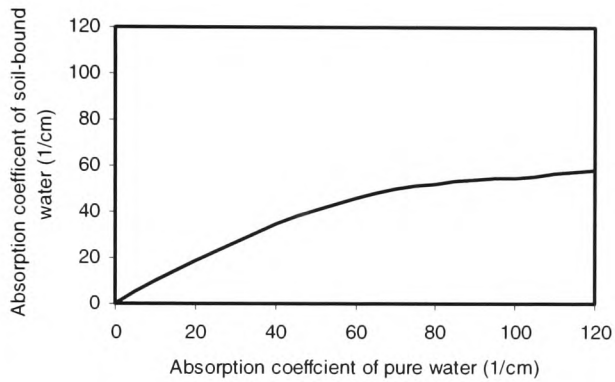


Figure 4.21 The relationship between the spectral absorption coefficients of soil bound water and pure water

Figure 4.21 shows the relationship between the spectral absorption coefficient for pure water and the coefficient for soil-bound water derived from inversion of the model. The figure shows that the relationship is not dependent on wavelength because the sharp increase in absorption in the water absorption bands is not represented in the data. The difference between the absorption coefficients increases with increasing absorption, with the soil-bound water absorption coefficient being considerable less than that for pure water at high absorption levels (i.e. within the absorption bands). Using the modified spectral absorption coefficients, excellent fits may be obtained between modelled and measured reflectances for a variety of soil moisture contents. This is illustrated in Figure 4.22 which is a plot of modelled versus measured soil reflectance for a number of different soils with varying water contents, using data measured by Mauser (1996). In Figure 4.22, the thickness of the water layer in the model was varied to achieve the best fit using a one-dimensional simulated annealing inversion. The water layer thickness is an arbitrary, empirical parameter that must be related to some more practical measure of soil moisture. Figure 4.23 shows the relationship between soil moisture measured by Mauser (1996) and the square root of the water layer thickness, l, of Equation 4.11. The different symbols in the Figure represent the different soil types studied and

although there is some scatter, there is a strong correlation throughout the range of soil moistures from dry to wet.

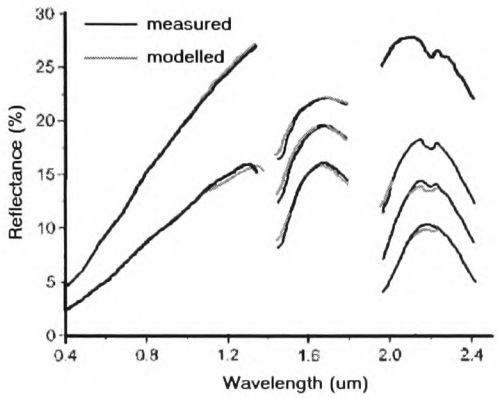


Figure 4.22 Measured and modelled reflectance spectra for soils with varying moisture contents, redrawn using data from Mauser (1996).

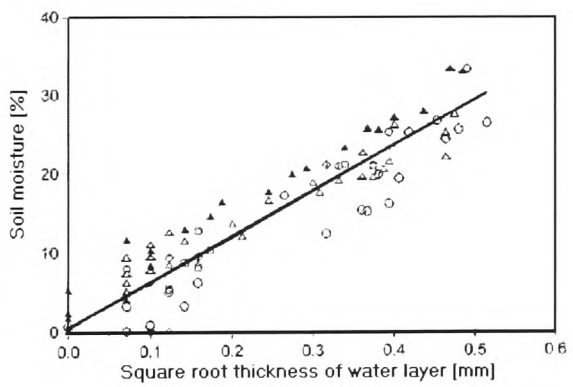


Figure 4.23 The relationship between measured soil water content soil water layer thickness as determined through optimisation of the model represented by Equation 4.11

The good performance of the model is encouraging, and its simple nature makes it easy to combine with SOILSPECT to form a model of soil spectral bidirectional reflectance as a function of soil moisture. In order to combine the model, the assumption has to be made that changing soil water content does not affect the soil bidirectional scattering characteristics, i.e. that the parameters b,b',c,c' and h are independent of soil moisture status. The seven parameters that control the combined model, SOILSPECT+ are given in Table 4-2.

Parameter name	*Symbol*
Single scattering albedo	ω
Soil roughness index	h
Bi-directional scattering coefficients	b
	b'
	c
	c'
Soil water layer	l

Table 4-2 The seven parameters of the SOILSPECT+ model.

Conclusion

The limited availability of directional data from remote sensing instruments makes BRDF model inversion using such datasets extremely difficult. Multi-sprectral data are much easier to obtain than multi-directional data, and a combined multi-directional and multi-spectral dataset would provide an excellent basis for model inversion.

The PROSPECT model is able to predict the spectral reflectance of leaves to a high degree of accuracy. It requires relatively short computing times, and has only five, wavelength-independent parameters. Moreover, it is readily inverted: in this chapter, excellent fits were obtained between optimized PROSPECT predictions and measured health leaf optical properties for species as diverse as millet and black spruce. Unfortunately, because of a lack of leaf biochemical composition data, it was not possible to determine whether the values of the five PROSPECT parameters retrieved by the optimization were accurate, but, the values appear plausible. By coupling PROSPECT with BRDF models, the wavelength dependent leaf optical properties that appear as parameters in the BRDF model may be replaced with the five wavelength independent PROSPECT parameters. Thus, only the soil reflectance term of the coupled model is a function of wavelength, and if this parameter too can be replaced by coupling a model of soil spectral reflectance to the PROSPECT-BRDF model, then the combined model would, in theory, be invertible against multi-directional, multi-spectral data.

The SOILSPECT model represents the bi-directional reflectance of bare soils with a reasonable degree of accuracy, using six parameters; four related to the anisotropic reflectance, and one related to the soil roughness, all of which may be regarded as wavelength independent. However, the sixth parameter, the single scattering albedo, ω, is not wavelength independent. It is this parameter which governs the spectral behaviour of soil reflectance and is dependent on both soil type and moisture status. The spectral dependence of ω complicates the inversion of SOILSPECT on spectral data, because its value is different in each measured spectral band, thus necessitating a large number of directional measurements in order to be able to achieve accurate model parameter estimation in the inversion.

The observed darkening of many surfaces when they become wet can be explained though an increase in the internal reflection at the liquid-air interface of the layer covering the surface combined with the strong absorption of water near 1400 nm and 1900 nm. A model describing these effects can successfully represent the change in reflectance of many soils with water content, provided the reflectance of each dry soil is known. Therefore, SOILSPECT may be modified to include equations describing the change in soil reflectance resulting in one extra input parameter (the water layer thickness). Using the combined model, a look-up table for the spectral dependence of the single scattering albedo, ω, need only contain entries for a variety of dry soils. Many BRDF models, including the the SAIL model used in this study, assume Lambertian soil reflectance, and in this case the combined soil reflectance model reduces to the look-up table of ω for dry soils and a model to describe the variation of ω with soil moisture content.

Chapter 5

Overview of SVAT Modelling

It is clear that extracting canopy parameters from remotely sensed data is not an easy task. There are significant problems with the use of all remote sensing techniques for this purpose, but the synergistic use of different techniques in combination may yield better results that the use of individual techniques alone (Chapter 3). The complex three-dimensional structure of plant canopies and the extremely heterogeneous nature of the Earth's surface make modelling the surface reflectance or interpreting reflections from the surface extremely difficult. Previous studies using optical data have concentrated either on the use of vegetation indices, or on attempting to invert reflectance models. Vegetation index datasets are plentiful, but suffer from signal saturation and interference from changes in background reflectance. The inversion of complex reflectance models is extremely difficult, but remains the only way of utilising optical data for many vegetation models, since these tend to be too complex to optimize over the long time steps at which RS data are available. Active microwave techniques appear to promise good results for the determination of biomass and LAI, but studies are currently suffering from a lack of data. Few studies have been made using radar remote sensing over crops. Signal saturation still seems to be a limiting factor for the study of dense forest canopies. In an analogous way to Chapter 2, which introduced the main concepts of remote sensing to readers unfamiliar with the subject, this Chapter is intended to introduce some of the basic concepts of vegetation modelling to readers, perhaps from a remote sensing background, unfamiliar with them.

5.1. MODELLING CONCEPTS

Vegetation models vary greatly in their design and complexity depending upon the application. In principle attempts to reproduce plant growth processes in order to make predictions about the future state of vegetation under specified climatic conditions. For validation purposes, it is usually necessary for models to predict variables that are measurable either in the field or by other means, for example remote sensing. Mechanistic models attempt to represent the various processes underlying plant growth and function in some detail. Empirical models simplify the representation of some or all of these processes.

Figure 5.1 shows, in schematic form, the functionality required by a mechanistic vegetation model. Plant metabolism occurs through photosynthesis in which light energy absorbed by leaves is used to convert CO_2 absorbed from the atmosphere into starch and sugars. Water and other nutrients required by photosynthesis are usually absorbed from the soil though root systems. Fluxes of gasses including CO_2 and water vapour diffuse into and out of leaves through pores, known as stomata. CO_2 from the atmosphere diffuses into leaves though the stomata. Stomata may open or close depending upon environmental and physiological factors. O_2 produced by photosynthesis diffuses out of the leaves through the stomata, together with water vapour. The loss of water vapour, and hence latent heat from leaves allows the plant to loose heat. However, in times of water shortage, plants may close their stomata to prevent water loss endangering health, through tissue damage or cavitation. The rate at which water vapour can diffuse through leaves also depends on atmospheric conditions such as the humidity and temperature of the air, and the aerodynamic resistance of the leaves, which is related to leaf shape and wind speed within the canopy.

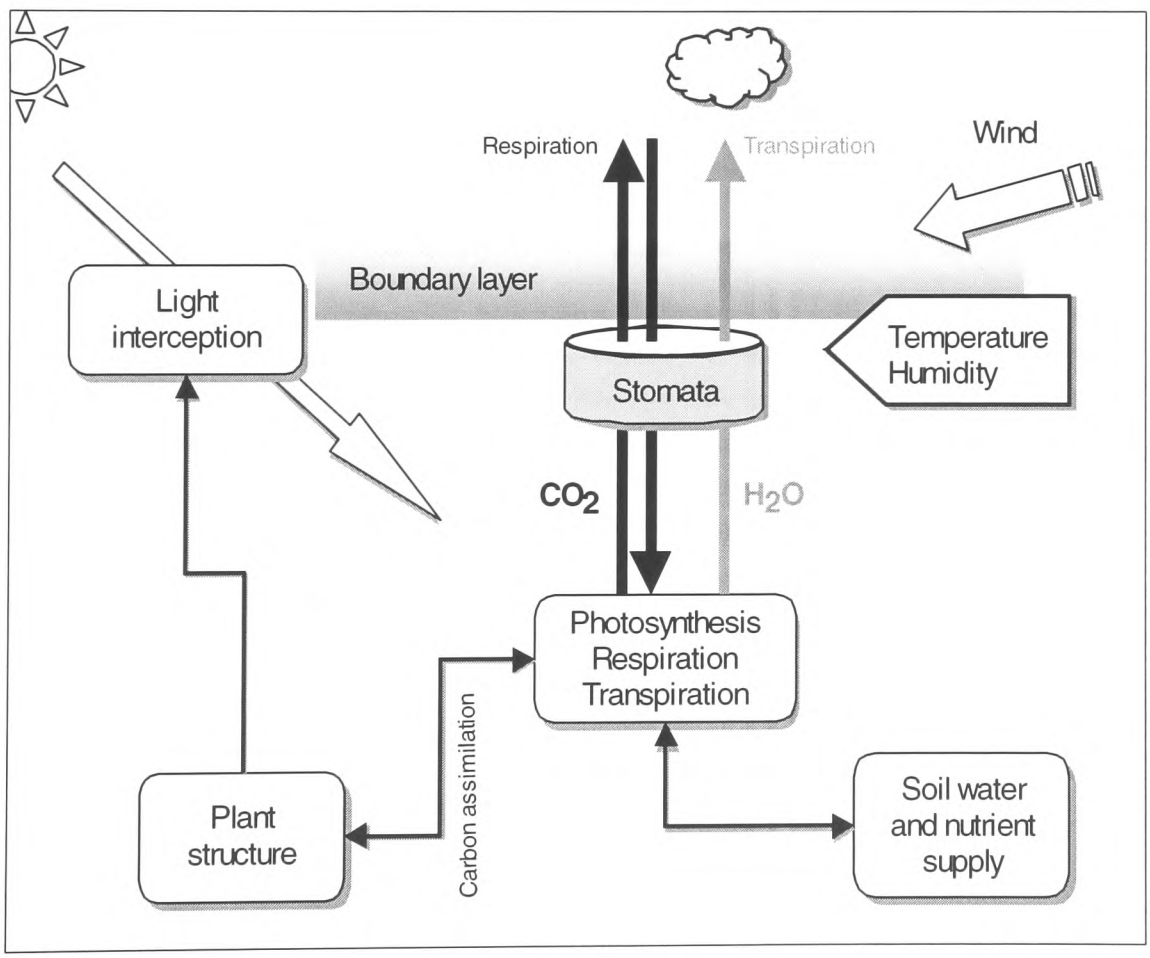


Figure 5.1 A simplified schematic diagram of the functioning of a plant.

Model Complexity

In the absence of water and nutrient stress, which is frequently the case for cultivars, light absorption is so important that empirical models relating light absorption directly to growth though a light use efficiency parameter ε have been very successful (Monteith 1977). More complex models attempt to model the processes of photosynthesis and respiration explicitly. Photosynthesis is an extremely complex process and two distinct forms (C_3, C_4) exist which differ in their reaction pathways. Many empirical models for photosynthesis have been developed, and whilst these models still dominate for C_4 photosynthesis, in recent years mechanistic C_3 photosynthesis models have been developed. The most successful of these models is undoubtedly the "Farquhar model" (Farquhar, von Caemmerer and Berry 1980), which models photosynthesis as a functions of enzyme activity, leaf internal CO_2 concentration, temperature and the electron transport within leaves.

A Question of Scale

Measurement of the physiological and optical characteristics of vegetation tend, by necessity, to be measured at the leaf scale. Predicting canopy scale fluxes from such measurements requires *bottom-up* scaling, by summing the leaf scale data over all the leaves in the canopy (Jarvis 1993). Unfortunately, this scaling requires knowledge of the variation of parameters throughout the canopy, which makes model parameterization difficult. In the alternative *top-down* approach, bulk parameters are determined from canopy-scale measurements, and used to characterize a "big leaf" model in which the canopy is treated as if it were a single, giant, leaf.

Canopy-scale variables cannot be treated as simple averages or totals of the values for individual leaves. There are two main reasons for this. Firstly, leaves do not act independently but interact with other leaves, and the surrounding atmosphere, modifying their own and their neighbours environment – this introduces feed-back effects. Secondly, processes defined by leaf-scale parameters respond in a non-linear manner to changes in environmental variables. A classic example of this is photosynthesis: assimilation rate saturates at high light levels, whereas light levels vary strongly with depth into the canopy. Thus the summation of leaf-scale parameters must be weighted by the actual leaf-scale flux (McNaughton 1994).

5.2. Big Leaf v Multi-Layer Models

The canopies of many vegetation species, notably forests, are complex three dimensional structures. Within the canopy air space the environmental conditions are not constant. The variation of light within the canopy is the most important factor. It is determined by the distribution of foliage (including branches, stems etc) and how light is intercepted by this foliage throughout the upper hemisphere. The distribution of foliage may be very complex indeed with clumping occurring at the leaf, twig and branch scales. Typically near the bottom of a dense forest canopy the light levels may

be a fraction of the level just above the canopy. Photosynthesis is a strong function of intercepted (and therefore incident) light and therefore may vary considerably with position within the canopy. However, there are other complex factors which also affect the rate of photosynthesis at each location in the canopy such as stomatal conductance, temperature, and factors which may be controlled by the plant itself, such as distribution of leaf area, nutrients and photosynthetic capacity.

Other important environmental parameters also vary with position (most importantly height) within the canopy. The wind speed may vary significantly, which affects the boundary layer conductance, humidity and air temperature. These in turn affect the stomatal conductance, which controls transpiration and photosynthesis. The aerodynamic roughness of a canopy determines how well the canopy air space is mixed with air above the canopy. This has important implications for flux measuring techniques such as eddy covariance. Rough canopies such as forests are typically well coupled to the atmosphere through turbulent mixing of the canopy air. The leaf temperature of such well mixed canopies is usually relatively constant with height and closely related to the air temperature. However, for relatively decoupled canopies significant differences between leaf and air temperature may exist.

The temperature of a plant, at any instant is determined by the energy balance equation:

$$R_n = M + C + \lambda_w E + G \tag{5.1}$$

Where: R_n is the net radiation flux density (heat gain from radiation, longwave + shortwave)
M is the net heat stored in biochemical reactions
C is the net sensible heat loss.
λ_w is the latent heat of vaporization of water
E is the flux of water vapour per unit area (evaporation rate)
G is the flux of heat by conduction

The net radiation is the most important term in Equation 5.1 because it drives many of the other fluxes. The value of R_n for a given incident radiation varies as a function of leaf temperature, and consequently as a function of factors affecting leaf temperature such as windspeed, humidity and leaf boundary layer resistance. This follows because the net radiation gain is the sum of all incoming fluxes minus all outgoing radiation, therefore:

$$R_n = R_{absorbed} - \varepsilon_s \sigma T_s^4 \tag{5.2}$$

Where: T_s is the temperature of the surface in Kelvin
ε_s is the emissivity of the surface
σ is the Stefan-Boltzmann constant.

This poses a difficulty in the calculation of leaf temperature T_s because C is a function of leaf temperature, ε is a function of leaf temperature through its dependence on stomatal conductance and saturation vapour pressure deficit, and R_n itself is a function of leaf tempertaure. Iterative computing

techniques can be used to determine T_s, however, by using the concept of isothermal net radiation (R_{iso}) (Monteith and Unsworth 1990) an analytical solution may be found:

Define:
$$R_{iso} = R_{absorbed} - \varepsilon_s \sigma T_a^4 \tag{5.3}$$

Substituting 5.5 into 5.4 and setting $T_s = T_a + \Delta T$

$$R_{iso} \approx R_n + 4\varepsilon_s \sigma T_a^3 (\Delta T) \tag{5.4}$$

Define a "conductance" to radiative heat transfer g_R as:

$$g_R = \left(\frac{4\varepsilon_s \sigma T_a^3}{\rho c_p} \right) \tag{5.5}$$

Then

$$R_n = R_{iso} - g_R \rho c_p (T_s - T_a) \tag{5.6}$$

Using 5.8, and 5.1, 5.4 may be solved (Monteith and Unsworth 1990) to give:

$$T_l - T_a = \frac{r_{HR}(r_{aw} + r_{sw})\gamma R_{iso}}{\rho_a c_p [\gamma(r_{aw} + r_{sw}) + \Delta_v r_{HR}]} - \frac{r_{HR} D_v}{\gamma(r_{aw} + r_{sw}) + \Delta_v r_{HR}} \tag{5.7}$$

Where:
- T_l is leaf temperature.
- T_a is air temperature.
- r_{HR} is resistance to heat transfer by radiation = $1/g_R$.
- r_{aw} is leaf boundary layer resistance to water vapour transport
- r_{sw} is leaf stomatal conductance to water vapour transport
- γ is the psychrometer constant
- Δ_v is the rate of change of saturation vapour pressure with temperature
- Dv is saturation vapour pressure deficit
- ρ_a is the density of dry air
- c_p is the specific heat capacity of air

However, when stomatal conductance is modelled as function of temperature, and leaf temperature in turn modelled as function of stomatal conductance, an iterative procedure is still required.

Modelling these non-linear relationships within canopy variations is a complex task. Many modellers have chosen to ignore the effects of a finite canopy height altogether by assuming that the canopy may be represented as a single layer of negligible height. These are the so called "big-leaf" models, first proposed by Monteith (1977) in which the canopy is thought of as being one giant leaf with a single associated value for photosynthetic rate and for stomatal conductance. Big leaf models are

widespread because they are far simpler to construct and require far few parameters than multi-layer models.

Multi-layer models such as MAESTRO (Wang and Jarvis 1990) are much rarer. They attempt to divide the canopy into layers or sub-volumes as different position within the canopy, within which canopy parameters are assumed to be constant. In order to calculate canopy scale variables, the results from each sub-volume must be weighted according to their relative amount in each layer or subvolume before being summed. Model calculations are typically made on a ground area or leaf area basis, in which case the contribution from each sub-volume must be weighted according to its fractional contribution to the total leaf area or total ground area.

Detailed mechanistic SVAT models such as MAESTRO, involve many parameters, and require detailed meteorological data at high temporal frequency. It is not feasible to use detailed mechanistic models of vegetation to calculate carbon balances over large areas because the inherent heterogeneity precludes their parameterization. Even if homogeneous areas within the landscape can be identified, it is not usually possible to specify parameter values for theses areas, nor provide the necessary detailed meteorological data. Consequently for large spatial scales simple models are required.

5.3. LIGHT USE EFFICIENCY

Light is the driving variable for all plant function (although respiration can occur in the dark, this results from earlier activities by the plant in light), therefore modelling light absorption by the plant is of the utmost importance. Monteith (1977) proposed that dry matter production is linearly related to the amount of PAR absorbed by vegetation:

$$P_p = \varepsilon Q_a \tag{5.8}$$

Where: P_p is the Net Primary Production (NPP)
Q_a is the absorbed PAR (APAR)
ε is the light use efficiency
(Note that although efficiency is strictly dimensionless. e usually has units of (g MJ^{-1})
P and Q_a are integrals over a sufficiently long period.

Where NPP is defined as:

$$P_p = G_p - R_a \tag{5.9}$$

Where: G_p is the gross primary production and is equal to he net photosynthesis.
R_a is the autotrophic respiration of the above and below ground biomass.

Much work has been carried out to discover whether ε is constant throughout the growing season. A relationship between dry-matter production and absorbed light seems to be obvious, because photosynthesis is a quantum dependent process. However there has been significant debate concerning this issue in the press (Demetriades-Shah *et al.* 1992 and 1994; Kiniry 1994; Monteith 1994).

Environmental effects on ε

Water stress

Over short time scales, high evaporative demand is likely to have an immediate effect on light use, through stomatal closure. Water stress over longer time periods is likely to affect plant growth and development processes, which may in turn lead to a reductions in dry matter production. This can occur though a reduction in leaf growth, or an increase in leaf fall which reduces the plant's ability to intercept PAR.

Temperature

The effects of temperature stress mainly act in the short term. Very low temperatures inhibit the photosynthetic apparatus, which may take several days to recover after warming. High temperatures inactivate enzyme reactions, but recovery is often rapid provided severe damage was not incurred. Similarly to water stress, long term temperature stress can lead to reduced growth and increased leaf fall.

Nutrient stress

Nutrient stresses operate directly on photosynthesis thorough effects on leaf enzymes such as Rubisco. These effects tend to be on the time scale of weeks to seasons. There is evidence that high tissue nitrogen levels increase the radiation use efficiency of crops (Sinclair and Shiraiwa 1993), and Dewar (1995) postulated that this results from increased leaf growth as opposed to increased photosynthetic capacity. Studies from forests are rather less numerous; but Wang, Jarvis and Taylor (1991) in a study of Sitka spruce found that nitrogen fertilisation increased ε by 30%, relative to unfertilised stands.

The importance of applying corrections for environmental conditions was illustrated in a study by McMurtrie *et al.* (1994), where data from several years for pine stands in four difference countries were analysed. The authors used the BIOMASS model (McMurtire *et al.* 1990) to predict "effective" values for absorbed PAR based on the assumption that not all the PAR absorbed by the canopy on a given day is utilized for photosynthesis if other conditions exist that prevent CO_2 assimilation. Potential photosynthetic gain is modified on the basis of several environmental factors according to:

$$G_p = \varepsilon \sum Q_a f_w f_D f_T \tag{5.10}$$

Where: G_p is the annual gross primary productivity.
f_x are modifying factors for a given day resulting from soil water deficit (w), saturation vapour pressure deficit of the air (D) and low temperature (T).

The modifying factor f_w was set to unity for high water levels and as a linear function of soil water below this critical level; f_D was given by a linear function of daytime VPD; f_T was set to small values if there were frost conditions on a given day. Figure 5.2 shows simulated GPP plotted against uncorrected APAR: considerable scatter is visible. However, a plot of GPP against "effective" APAR derived using Equation 5.12 shows a much better correlation (Figure 5.3.)

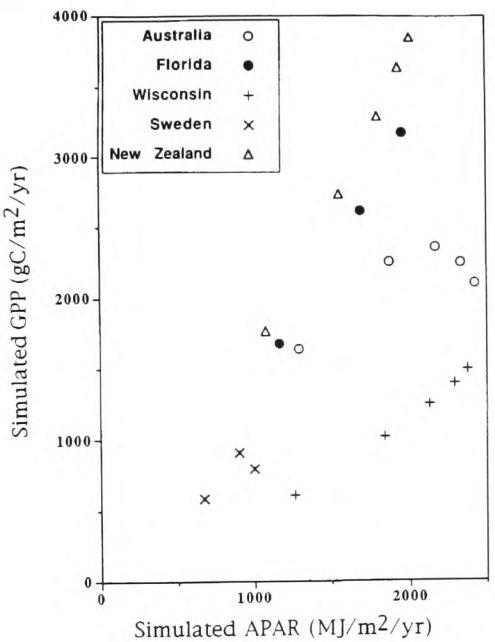


Figure 5.2 Simualted gross primary production (GPP), uncorrected for environmental effects on ε, for pine stands in several countires, plotted against simulated APAR calculated using a radiation absorption model, (redrawn from McMurtrie *et al.* 1994)

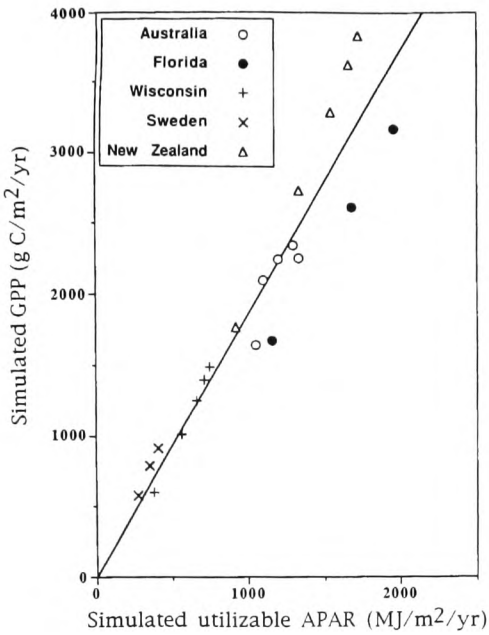


Figure 5.3 Simulated GPP corrected for environmetal effects on ε using Equation 5.12 (redrawn from McMurtrie *et al.* 1994)

Study	*Plants*	ε g dry matter MJ^{-1}
Monteith, 1977	sugar beet, barley apples, potatoes	1.4
Cannel *et al.*, 1987:	willow (above ground)	1.4
Russell *et al.*, 1989:	various intensely managed field crops	2.8
Landsberg and Wright, 1989	populus (first year of growth)[1]	0.32
	(second year)[1]	1.4
Jarvis and Leverenz, 1983	warm area deciduous forests	0.15
	cool-area evergreens	0.78
Wang *et al.*, 1981	Sitka spruce (fertilized)	0.95
	(unfertilized)	0.75
Saldarriaga and Luxmoore, 1991	tropical rain forest (young stand)[2]	0.35
	(mature stand 10-50 years)	0.2

Table 5.1 Values of light use efficiency ε obtained by various studies.

[1] The small value of ε in the first year, relative to the second year was attributed to a large proportion of the dry matter produced being used for root development in the first year.

[2] The very low values of ε reflect the very low mineral nutrition at the field sites in Columbia and Venezuela.

Physiological basis for ε

The rate of change of photosynthesis with PPFD is highly non-linear, however, if the canopy is closed much of the foliage will not be PPFD saturated. Under these circumstances the photosynthesis of the canopy as a whole is light-limited and the relationship of canopy photosynthesis to intercepted light tends toward linearity (Charles-Edwards 1982).

Any non-linearities in the relationship are likely to be less apparent over longer time periods. Wang *et al.* (1992) used the BIOMASS model (McMurtrie *et al.* 1990) to study the relationship between daily photosynthesis and APAR. They showed that there was considerable scatter between daily photosynthesis and APAR, but this variability disappeared when annual values of photosynthesis and APAR were used. In another study, Jarvis and Leverenz (1983) showed that photosynthesis of a Sitka spruce canopy was linearly related to PPFD, largely as a consequence of shoot and canopy structure.

Analysis of the light use efficiency should not be based on photosynthesis alone - any physiological interpretation must also consider respiratory processes. It is commonly observed that photosynthesis exhibits a saturating response to increasing leaf nitrogen content, however, the response of respiration is almost linear. This implies that there is a value of plant nitrogen content where net primary production (NPP) has a maximum value. Dewar (1995), using these ideas, formulated a model of NPP using the dependence of canopy photosynthesis and respiration on nitrogen content. By assuming that plants optimize nitrogen distribution to maximize NPP, Dewar (1995) performed this optimization in two stages:

1. Maximize gross canopy photosynthesis with respect to the relative nitrogen distribution within the canopy.
2. Maximize NPP with respect to the relative nitrogen distribution within the canopy for a given leaf area index.

By assuming that growth respiration is proportional to NPP (Thornley and Johnson 1990), and that maintenance respiration is strongly correlated with tissue nitrogen content, Dewar (1995) showed that the relative optimal distribution of nitrogen followed the light extinction profile through the canopy, and that under these conditions the canopy acts as a big leaf. According to the model the maximum value for NPP is given by:

$$W_{\max} = I_{day}\alpha Y_g (1-\lambda)^2 \qquad (5.11)$$

Where: $W_{\max}$ is the maximum daily net primary production (NPP)
I_{day} is daily PAR intercepted by the canopy
α is the quantum yield (the amount of carbon fixed by photosynthesis per joule of PAR absorbed)
Y_g is the biosynthetic efficiency of the conversion of sugars formed during

photosynthesis to structural dry matter during growth.

λ is a parameter that characterized the relative nitrogen sensitivities of respiration and light saturated photosynthesis.

λ is given by:

$$\lambda = f\left(N, r, \lambda^*, T, \beta, C_a, h, k_x\right) \tag{5.12}$$

Where: N is the tissue nitrogen content per unit area

r is the maintenance respiration per unit N

λ^* is the nitrogen partitioning of the plant between leaves, roots and other tissues.

T is daily average temperature

β is the ratio of leaf internal CO_2 concentration (C_i) to C_a.

C_a is the atmospheric CO_2 concentration

h is the daylength

k_x is a species specific carboxylation coefficient that determines the rate at which light-saturated leaf photosynthesis increases with increasing leaf nitrogen content.

The light use efficiency ε is given by:

$$\varepsilon = \alpha Y_g \left(1 - \lambda\right)^2 \tag{5.13}$$

and is independent of incident PAR. Equation 5.11 describes how the value of ε is reduced from its theoretical maximum, because of the effect of growth (Y_g) and maintenance ($1 - \lambda$) respiration and by the quantum yield, α.

Dewar (1995) suggested that this analysis offers a possible physiological interpretation for the linear relationship between NPP and daily APAR and also identified a number of physiological and environmental factors that may influence ε.

Parameter	*Effect on ε*
Quantum yield (α)	Increasing
Biosynthetic efficiency (Y_g)	Increasing
Maintenance respiration (r) [1]	Decreasing
Nitrogen partitioning (λ^*)[2]	Decreasing
Carboxylation coefficient (k_x)[3]	Increasing
Ration of internal to ambient CO_2 conc. (β)[4]	Increasing
Ambient CO_2 concentration (C_a)	Increasing
Daylength (h)	Increasing

Table 5.2 The effect of an increase in various physiological and environmental parameters on ε, (from Dewar 1995).

[1] Parameter increases with increasing temperature.

[2] ε decreases with increasing partitioning of nitrogen into root and other non-foliage tissues.

[3] Parameter increases with increasing temperature.

[4] Parameter decreases linearly with increasing saturation vapour pressure deficit (Monteith 1995).

A further important implication of this analysis, embodied by Equation 5.12, is that the value of ε is independent of nitrogen supply. This is a direct consequence of the fact that NPP is maximized with respect to nitrogen supply and is limited only by light interception (Dewar 1995). This, in turn, implies that nitrogen fertilization increases NPP through an increase in light interception caused by enhanced leaf growth, rather than through an increase in the light-saturated photosynthetic capacity of individual leaves.

5.4. ACCLIMATION TO LIGHT

For multi-layer models in order to calculate canopy-scale processes it is necessary to integrate leaf-scale processes through the canopy. Unfortunately, numerous studies have indicated that photosynthetic and stomatal characteristics of leaves show phenotypic plasticity and adjust to the prevailing environmental conditions, e.g. temperature (Björkman 1981a), CO_2 (Sage, Sharkey and Seeman 1989) and radiation (Björkman 1981b). Phenotypic plasticity may be associated with physiological modifications during plant developments leading to variations in leaf thickness or stomatal density (Weyers, Lawson and Peng 1997). Over time scales of days, changes may also occur in the amounts of photosynthetic enzymes (Björkman *et al.* 1972), and on even shorter times scales the activity of these enzymes (Sage *et al.* 1989).

The resources available to a plant are limited. Clearly, it would be advantageous to use these limited resources in the most efficient way. Measurements for forests canopies show that the photosynthetic capacity, maximum stomatal conductance, base respiration rate and leaf nitrogen content expressed on an area basis, all decrease with depth of foliage (Lewandowska, Hart and Jarvis 1977; Field 1983). Moreover, experimental research (e.g. Evans 1988) has shown that the photosynthetic system is usually well adjusted to the light environment, and does not in general carry any excess capacity. Plants are able to redistribute resources over a period of five to ten days to exploit changes in irradiance, and are thus able to acclimate to the light environment. The strength of this acclimation apparently varies between species and over time (Oberbauer and Strain 1986) but further comprehensive studies are required in order to determine whether acclimation is widespread amongst plant communities.

Modelling Photosynthesis and Transpiration

The question of acclimation of photosynthesis to the local light environment has been considered by several authors (Kruijt, Ongeri and Jarvis 1997; Sellers *et al.* 1992; McNaughton 1994; Dewar 1995). Here, the physiological relationships between photosynthesis and both transpiration and PAR, and the implication of these relationships for vegetation modelling, are described, following the approach of Kuijt, Ongeri and Jarvis (1997).

The Farquhar and von Caemmerer (1982) photosynthesis model expresses CO_2 assimilation rate, A as being either limited by electron transport rate, J, or, if PAR is saturating, by rubisco activity which determines a maximum carboxylation rate V_{cmax}. In addition, A is a function of temperature, T and internal CO_2 concentration C_i (see Appendix B for mathematical details of the model). Therefore:

$$A = \min\begin{Bmatrix} V_{c\max} \cdot f(T, C_i) \\ J \cdot g(T, C_i) \end{Bmatrix} - R_d \qquad (5.14)$$

Where: $f()$ & $g()$ are different functions of temperature and internal CO_2 concentration.
R_d is the dark respiration rate, which is an exponential function of T.

J is determined for a PAR response curve, typically of the form (Evans and Farquhar 1991):

$$J = \frac{\alpha I_a + J_{\max} - \sqrt{\left((\alpha I_a + J_{\max})^2 - 4\Theta\alpha I_a J_{\max}\right)}}{2\Theta} \qquad (5.15)$$

Where: I_a is the absorbed PAR
J_{max} is the maximum electron transport rate
Θ is a curvature factor (Leverentz *et al.* 1990)
α is the initial slope of the light response curve (i.e. the quantum yield on an absorbed PAR basis)

Now since $A = g_s(C_a - C_i)$, substituting for C_i in Equation 5.1 enables that equation to be solved in terms of the ambient CO_2 concentration C_a and the stomatal conductance for photosynthesis g_s.

The modelling of stomatal conductance is still the subject of considerable debate, with one of the issues being whether the ratio C_i/C_a should be conserved at a near constant value. Stomata tend to open as C_i decreases but the sensitivity is strongly dependent on species and environmental conditions. The value of C_i is maintained at surprisingly constant levels in most species over a wide range of conditions (Wang *et al.* 1979), which would occur if stomatal conductance varied in proportion to the assimilation rate. However, mechanistic evidence for this hypothesis is lacking and stomata respond to CO_2 in both light and dark, so the response cannot depend solely on photosynthesis. It seems likely that the observed behaviour is as a result of the close matching of stomatal conductance and assimilation rate. A wholly satisfactory mechanistic model has yet to be developed, consequently empirical stomatal conductance models are employed in most studies. Jarvis (1976) proposed an empirical approach where g_s is expressed as follows:

$$g_s = g_0 \cdot f_1(I_a) \cdot f_2(D_v) \cdot f_3(T) \cdot f_4(\Psi_l) \cdots \qquad (5.16)$$

Where: f_n are individual functions formed by results obtained from controlled studies
D_v is the saturation vapour pressure deficit of the atmosphere
T is temperature
ψ_l is the leaf water potential
g_0 is a reference value for given environmental conditions.

Consider a formulation of the Jarvis approach as follows:

$$g_s = \frac{g_{sw}}{1.6}$$

$$g_s = \frac{1}{1.6} \frac{g_{sw\max} \cdot \alpha(I_a + \beta)}{g_{sw\max} + \alpha(I_a + \beta) \cdot h(T, D_v, C_a)} \tag{5.17}$$

Where: g_s and g_{smax} are the maximum and actual values of the stomatal conductance for water vapour
α and β are the initial slope and intercept of the PAR response respectively
h is some function of T, D_v and C_a

Equation 5.4 does not enforce a constant ratio of C_i/C_a, however, in practice the responses of photosynthesis and stomatal conductance to PAR and CO_2 are such that this ratio is kept fairly constant.

Transpiration can be modelled according to the Penman-Monteith equation (Monteith and Unsworth 1990) as modified by Jarvis and McNaughton (1986):

$$E = c_1(1-\Omega)g_{sw}D_v + c_2\Omega Q_a \tag{5.18}$$

Where: c_1 & c_2 are temperature-dependent coefficients
Ω is a decoupling factor which is a function of the ration of the g_{sw} and g_b, the boundary layer conductance of the leaves.
Q_a is the available energy (absorbed radiation in all wavelengths)

The above equations for photosynthesis and transpiration are non-linear with respect to I_a. This means that rigourous upscaling of V_{cmax}, J_{max}, g_{swmax} and R_d from leaf scale to canopy scale necessitate the weighting of each of these values with the flux predicted at each leaf, and averaging over all the leaves in the canopy. Even when these parameters are constant throughout the canopy, this results in an intractable integration (Sellers *et al.* 1992). If however, these parameters reflect full acclimation to the light environment, then Sellers *et al.* (1992) show that the integration becomes possible.

For a fully acclimated canopy the values of V_{cmax}, J_{max} and g_{swmax} at any point in the canopy can be calculated from the known values at any given reference location using:

$$B_e = B_0 \frac{\bar{I}_{ae}}{\bar{I}_{a0}} = B_0 p_e \tag{5.19}$$

Where: B denotes any one of the parameters
p_e is the fractional average absorbed PAR for a given canopy element relative to a reference element
The overbars denote an average over a sufficiently long acclimation period. The subscript e refers to a given canopy element and the subscript 0 refers to the reference location that may be anywhere in the canopy.

Although the distribution of nitrogen is not explicit in the assumption of full acclimation, V_{cmax}, J_{max} and g_{swmax} are strong functions of photosynthetic enzymes, and since leaf nitrogen is largely contain within such enzymes, it is reasonable to expect that the distribution of nitrogen will mimic the distribution of the other parameters. Ryan (1995) showed that respiration is probably a function of leaf nitrogen, and therefore Equation 5.6 is likely to hold for respiration.

If the distribution of absorbed PAR is constant over time, if the ratio C_i/C_a is constant,[3] and if β is small, then following substitution of Equation 5.6 into Equations 5.1, 5.2 and 5.4, it can be shown that photosynthesis is a linear function of p_e (Kruijt, Ongeri and Jarvis 1997). By defining q_e as the ratio of the average available energy for a given canopy element to that for the reference element (in an analogous manner to p_e) and substituting into Equation (5.5), then transpiration can be made a linear function of the relative absorbed radiation through the canopy, provided Ω is constant through the canopy (i.e. assuming the g_{sw} and g_b decrease in a similar manner relative to canopy position).

Using the above equations, the instantaneous values of A, E and g_s depend upon:

1. The value of , V_{cmax}, J_{max}, g_{swmax}, R_d and absorbed PAR in the reference location.
2. A function of T, D_v and C_a.
3. The ratio of the averaged absorbed PAR and available energy, p_e and q_e.

For canopies that are well coupled to the atmosphere such as forests, T, D_v and C_a are relatively constant with depth in the canopy, but p_e and q_e, of course, vary strongly. For such a canopy, the integration of photosynthesis for all leaves per unit ground area reduces to the integration of p_e and q_e over all canopy elements (Kruijt, Ongeri and Jarvis 1997):

$$\Pi = \int_{canopy} p_e$$
$$\Xi = \int_{canopy} q_e \tag{5.20}$$

Where: Π is the canopy PAR absorption relative to absorption by leaves

[3] C_i/C_a is assumed to be constant through a similar dependence of g_s and A, on PAR.

at the reference location.

Ξ is the relative canopy absorption of available energy.

Photosynthesis and transpiration may then be scaled as follows:

$$A_c = \left[\min \left\{ \begin{matrix} V_{c\max 0} \cdot f(T, C_i) \\ J(J_{\max 0}, I_{a0}) \cdot g(T, C_i) \end{matrix} \right\} - R_{d0} \right] \cdot \int_{canopy} p_e L_e \qquad (5.21)$$

$$A_c = A_0 \frac{\bar{I}_{ac}}{\bar{I}_{a0}} = A_0 \Pi$$

$$g_{sc} = g_{s0} \Pi$$

$$E_c = c_1 (1 - \Omega) g_{sw0} D_v \cdot \Pi + c_2 \Omega Q_{a0} \Xi \qquad (5.22)$$

Where: L_ε is the leaf area of a given element.
The subscript c refers to a bulk canopy value, and the subscript 0 refers to the reference location that may be anywhere in the canopy.

Note that A_c and E_c in Equations 5.8 and 5.9 are expressed on a ground area basis, whereas the fluxes at the reference location are expressed on a leaf area basis. This is accounted for by Π and Ξ having units of m^2 leaf/m^2 ground. Equations 5.8 and 5.9 form a big leaf model that may be parameterized using leaf scale measurements made at one location in the canopy.

The distribution of PAR in the above equations is assumed to be constant over time. However, because of variations in solar position, and the diffuse fraction of incident radiation, this assumption is likely to be untrue. For a given canopy element, p_e (and q_e) will deviate from its average value by an amount $\gamma_e = \frac{I_{ae}}{p_e I_{a0}}$. Kruijt, Ongeri and Jarvis (1997) examined the predictions may by a three-dimensional forest model (MAESTRO; Wang and Jarvis 1990) with a bulk canopy model (based on Equations 5.8 and 5.9) that assumes acclimation of photosynthesis, and examined the importance of the factor γ_e. MAESTRO was parameterized using Equation 5.6 to relate the values of the canopy parameter at a given location in the canopy to values at the top of the canopy (the reference location). Therefore the main difference between the two models was the explicit calculation of the distribution of absorbed radiation and leaf scale fluxes in MAESTRO, which enabled variation of γ_e along the PAR gradient. Using measured parameters and meteorological data for a stand of Sitka spruce (*Picea sitchensis*) in Scotland, they found that the prediction of photosynthesis by both models was in

extremely good correspondence, but that the bulk model tended to underestimate transpiration by between 5% and 10%, relative to MAESTRO. This was explained in part, by the stronger overall dependence of transpiration on g_{sw} and hence on I_a. The sensitivity of transpiration to the ratio g_{sw}/g_b is much stronger than that for photosynthesis. Although both conductances decrease with depth into the canopy, this is not always at the same rate, therefore Ω is not necessarily constant, undermining a basic assumption of the bulk model. Careful examination of the results also showed a diurnal trend in the difference between the two models to be evident, with the bulk model underestimating transpiration relative to MAESTRO in the early morning shortly after sunrise, and overestimating transpiration shortly before sunset. For large solar zenith angles, radiation reaching the bottom of the canopy is strongly attenuated because of the high probability of the light passing through other crowns. This results in I_a decreasing more rapidly with depth than for nadir illumination, and since the reference location was chosen to be at the top of the canopy, this explains why the bulk model tends to overestimate transpiration fluxes near sunset. A corresponding effect near sunrise was not observed because at these times of day there tended to be extensive cloud cover, and the fraction of diffuse radiation was large. A high proportion of diffuse radiation results in a less steep PAR gradient within the canopy, thus explaining the observed slight relative underestimation of transpiration by the bulk model.

The assumption of acclimation leads to very useful simplifications in the modelling of canopy scale fluxes. Once PAR at a reference location is known (through the use of a radiation transfer model), acclimation leads to greatly simplified photosynthesis and transpiration models that are effective single layer, big leaf models. However, the validity of the acclimation assumptions has yet to be proven through experiment, and important questions concerning the extent and species-specific nature of acclimation remain to be answered.

Mechanistic models of photosynthesis, such as those using the Farquhar model, usually have a time step of 60 minutes because these time steps are convenient for comparison and validation against eddy covariance measurements. The disadvantages of using hourly time steps are obvious; meteorological and incident radiation data are required at high frequency, with the inherent effects on computation time. Consequently, hourly time step models are usually run over time periods of no greater than 12 months, and are frequently used to examine the effects of environmental or physiological changes on photosynthesis.

When the intention is to run the model over time scales of years, or even centuries, the model time step tends, by necessity, to be longer. Longer time steps are also used with computationally intensive models where the representation of the photosynthetic process is perhaps of secondary importance to other areas of study. An example of this latter case might be a crop model where the aim is to study

the phenology and detailed development of the canopy in the form of tillers and leaves. In this case a simple empirical model of photosynthesis might be used with a 24 hour (daily) time step.

5.5. PARAMETERIZATION OF SVATS WITH REMOTELY SENSED DATA

Detailed models of photosynthesis require a very large number of parameters, most of which are impossible to measure using remote sensing (RS) techniques (Chapter 2). The time step of models is also a limiting factor for the use of remotely sensed data. Current RS techniques (both optical and microwave) can only detect changes in canopy attributes such as LAI over a period of weeks. In addition, the availability of optical RS scenes is further limited by the presence of cloud cover. This makes the regular parameterization of SVATs impossible.

As described in Chapter 2, the important vegetation attributes that can be reliably obtained from RS data are probably limited to APAR, LAI and biomass. The means that, with the exception of empirical models such as LUE models (Section 5.3), frequent meteorological data are required, which can only be obtained from weather stations situated in reasonably close proximity to the target vegetation. No one RS technique for parameter estimation is universally applicable, and all suffer from some form of disadvantage. For example, NDVI data, although widespread and readily available, suffer from signal saturation at moderate LAIs and are sensitive to changes in background reflectance. The synergistic use of RS data from different sources might not only reduce the problems associated with each technique individually, but increase the frequency with which data are available. It would seem advantageous, therefore, to develop a model that could make use of a variety of RS data, as and when these are available during the year.

Several studies of the coupling of remotely sensed data and vegetation models have been reported, varying in the complexity of the model used and the type of information derived from the remotely sensed data. Because of the difficulties involved with integrating RS data with small-scale models (Chapter 2), most of these studies have been at the regional or global scale. These studies are based on two approaches outlined below.

Inversion - parameterization

Either inversion techniques, or simple empirical relationships, are used to derive biophysical (and in some cases even environmental) variables from RS data. These variables are then used to parameterize the vegetation model, (Figure 5.4a). This *forcing strategy* has been widely used for both crops (e.g. Maas 1988) and natural vegetation (e.g. Running *et al.* 1989). Most studies have involved

the use of vegetation indices, however, empirical relationships such as vegetation indices, suffer from signal saturation and are therefore insensitive to changes in LAI above 2 to 4. BRDF models utilize directional information to give a more accurate estimation of the anisotropic reflectance from vegetation canopies and could theoretically be inverted to obtain canopy biophysical information. Unfortunately, inversion techniques exhibit sensitivity to noise, and detailed reflectance (BRDF) models that are able to model the reflectance of vegetation canopies accurately are extremely difficult to invert.

Validation - optimization

Examples of these approaches in the literature are far rarer; they consist of deriving surface radiative properties from a vegetation model and comparing them with RS data (Figure 5.4b). The vegetation model parameters are then optimized to find the best fit between the modelled and measured reflectances over time. In one of the few examples of a validation-type technique, Nemani and Running (1989) used NDVI data derived from AVHRR, and a forest ecosystem model (FOREST-BGC), to test the concept of equilibrium between LAI and water availability. The authors concluded that the combination of ecosystem simulation and remote sensing was able to provide insights into canopy dynamics and carbon partitioning

For detailed mechanistic models (such as those based on the Farquhar model of photosynthesis) that require an hourly time-step, the forcing strategy remains the only one available, because the number of variables and the temporal dynamics are such that computation effect involved precludes the use of the optimization-validation approach. The best that can be realistically achieved with these models is either a direct validation of the model (without optimization) or irregular parameterization via LAI estimates. For empirical or pseudo-empirical models, the validation-optimization approach provides an attractive alternative. In this technique, the vegetation model is parameterized with a base set of variables and run forward until a time when RS data are available. A surface "reflectance" model that is linked to the vegetation model is then used to predict the RS signature of the canopy, which can be compared against the measured data. Over a period of time, and several remote sensing *scenes*, the vegetation model can be optimized, or constrained, to fit the RS data (Figure 5.4b). The purpose is to find the trajectory of model variables (principally LAI) that best fits a series of observations distributed time. Remote sensing is used to detect changes in the canopy that occur over time, this will mainly be as a result of growth processes, and, depending on growth rate, changes are likely to be detectable on a time scale of weeks.

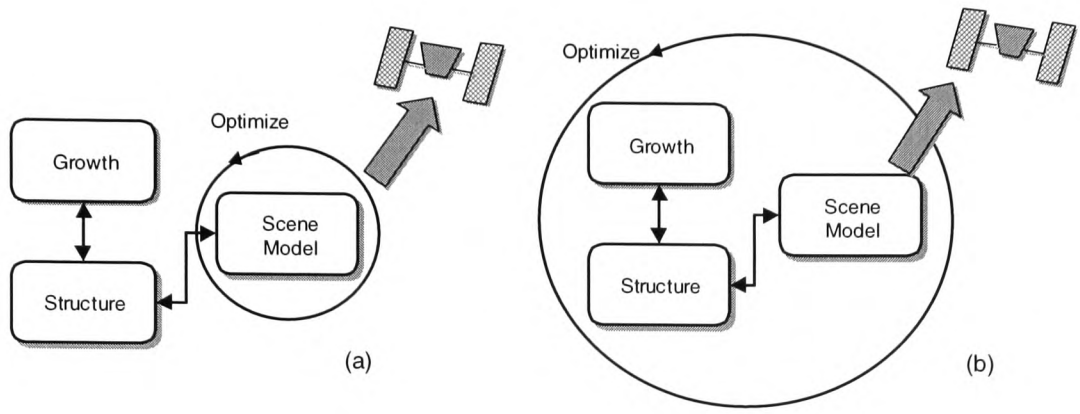


Figure 5.4 Schematic representation of the (a) inversion-parameterization and (b) the validation-optimization approach. The *scene* model represents some way of predicting the signal at the remote sensing instrument based on canopy biophysical parameters, in particular LAI or biomass. Typically the this takes the form of a BRDF model.

As discussed in Chapter 2, it is envisaged that BRDF data will form the main source of remotely sensed information (the *scene* model) for two reasons. Firstly, BRDF measurements contain significantly more information about the vegetation canopy than nadir measurements, and secondly, although promising results have been reported with radar techniques, particularly with the estimation of forest biomass, active microwave data are not yet readily available. The modelling of changes in canopy structure resulting from growth process predicted by a SVAT can be used as inputs to a BRDF model, creating a coupling. By coupling a BRDF model to a SVAT, not only may directional reflectances be predicted, but the BRDF model may also be used to calculate APAR, providing useful twin rôles.

The limitation of using simple empirical growth models is obviously a disadvantage, however, if the vegetation model were properly designed, it might be possible to use the validation-optimization approach described above, together with a daily time step empirical model, to re-calibrate more mechanistic models. In this way the model "changes mode" depending on the application, as is illustrated in Figure 5.5.

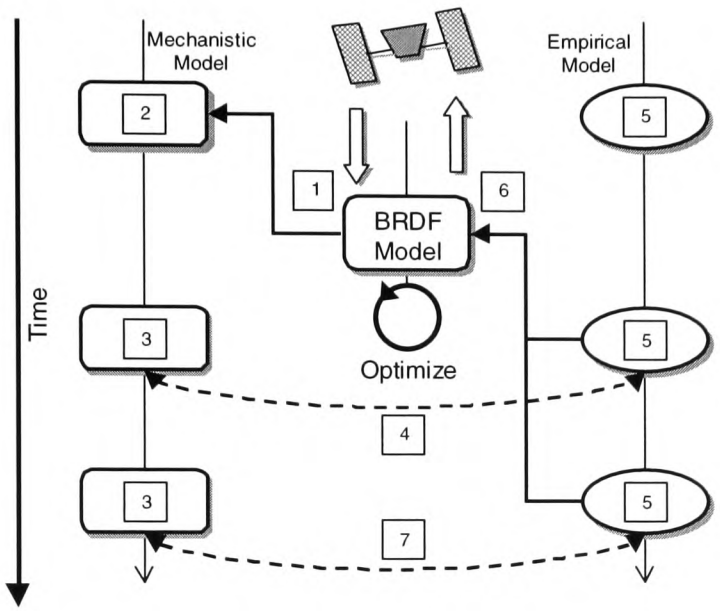


Figure 5.5 The steps involved with the inversion-parameterization and validation-optimizaton approaches for a coupled SVAT-BRDF model, showing the linkage between a complex mechanistic and a simple empirical vegetation model.
(1) Single date remotely sensed canopy directional reflectances can be used to invert a BRDF model to obtain estimate of canopy bio-physical variables such as LAI.
(2) These variables can then be used to parameterize a SVAT.
(3) Further remotely sensed measurements become available on other dates.
(4) Mechanistic model variables can be used to help parameterize the empirical model.
(5) The empirical model is run on a daily time step to predict canopy growth over the period. Growth predictions from the empirical model are used in a coupled BRDF model to predict canopy reflectances.
(6) These reflectances are then compared against remotely sensed measurements, and the model is optimized to produce the best fit trajectories of the model parameters over time.
(7) The optimized empirical model can be used to recalibrate the mechanistic SVAT.

In order to utilize measured data dispersed over time, such as those typically available from remote sensing, or from occasional, field campaigns, the optimization routine must form part of the model itself. Although this greatly adds to the complexity of the model design, it enables complete freedom in the choice of data inputs and allows for a more systematic calibration. The flexibility of the approach outlined in Figure 5.5 means that other remotely sensed data may be readily incorporated in to the modelling scheme. For example, the BRDF model can be used to predict NDVI values instead of directional reflectance in steps 5 and 6 of Figure 5.5, thus constraining the vegetation model against measured vegetation indices. Alternatively, if estimates of biomass (and hence LAI) are available from radar remote sensing, then the optimization step can proceed using the vegetation model alone,

without need for the coupled BRDF model. For maximum flexibility, and the maximum use of synergy between remote sensing derived information, the modelling scheme should be able to incorporate all of the above forms of data in an automated fashion.

5.6. CONCLUSION

It is clear from the above discussion that in order to make regular and routine use of remotely sensed data for vegetation modelling a new modelling strategy is required. This strategy should be able to combine remotely sensed data from a number of sources in a synergistic way. Radar remote sensing techniques, although offering promising results for biomass estimation, especially for forests where dense canopy preclude the use of optical techniques, suffer from limited data availability. Therefore, optical data will remain the primary source of RS information for the immediate future. The increasing availability of directional reflectance data enables the coupling of BRDF models to SVATs. The BRDF signal contains more information about canopy properties than vegetation index measurements, and is less prone to signal saturation. In addition the coupled BRDF model can be used to derive absorbed radiation estimates that are the key driving variable for canopy fluxes and growth. The traditional strategy for using BRDF data has been the inversion-parameterization approach, but the complexity of BRDF models and the large number of parameters they require make inversion extremely difficult. The frequency with which optical RS data are available during the growing season are usually restricted by the presence of cloud cover, which makes regular parameterization of models impossible. The alternative validation-optimization strategy presented here makes use of multi-temporal measured data corresponding to changes in canopy structure resulting from growth processes, to constrain vegetation model behaviour over time. Although, detailed mechanistic SVATs are usually too complex for this approach, they can be recalibrated through linkage to a simpler empirical growth model.

The fact that canopy scale variables cannot be treated as simple averages of the values for individual leaves leads to complexities in vegetation modelling. In general, there are two approaches to this problem. The big-leaf approach treats the canopy as if it were a single leaf with a single associated value for photosynthetic rate and stomatal conductance. The multi-layer approach divides the canopy into a number of layers, and canopy scale variables the results from each sub-volume are weighted according to their contribution and extent. Models usually fall into one of these categories, however, the modelling scheme shown in Figure 5.5 may encompass both a multi-layer mechanistic model and a big-leaf empirical model, therefore it is desirable that this flexibility is included in the model design.

Optimization clearly plays a vital rôle irrespective of whether the inversion-parameterization or validation-optimizaton approach is used. The next chapter discusses four different optimization techniques and compares their relative features and merits. Chapter 7 then goes on to discuss the application of these optimization algorithms to the inversion of BRDF models and in particular the SAIL model. Chapter 8 describes the design and construction of a coupled SVAT-BRDF model that contains the features outlined above.

Chapter 6

Optimization Techniques

Optical remote sensing instruments measure the radiation reflected or emitted from the Earth's surface. In order to obtain information useful to vegetation modellers, either model parameters are inferred from empirical relationships with reflectance such as vegetation indices, or models of surface reflectance must be inverted. BRDF models predict canopy reflectance based on a number of input parameters and inverting such a model involves taking the measured reflectance data and running the model backwards in an attempt to find the set of input parameters that would generated the measured reflectance. It is frequently impossible to rewrite the model in reverse form, and therefore the inversion problem reduces to an iterative process of making an initial guess for the model input parameters, comparing the output of the model with the measurements, and refining the initial guess. Such problems occur extremely frequently in most areas of science and engineering and this field of numerical research is generally known as optimization.

The problem of optimization can be expressed in the following way:

> Given a model described by the function $f(\boldsymbol{x})$,
>
> Where: $\boldsymbol{x}$ is an n dimensional vector of independent parameters
>
> Find $\boldsymbol{x}$ to minimise or maximise $f(\boldsymbol{x})$

The problems of minimisation and maximisation are trivially related and will henceforth be known simply as optimization. The goodness of fit of a function to a set of measured data, depends upon a number of factors: the accuracy of the model representation itself, the noise inherent in the measurements, and the effectiveness of the optimization algorithm in finding the global minimum amongst any local minima that may exist. The model $f(\boldsymbol{x})$ is often very complex and since the optimization is usually an iterative process, the quality of the solution (defined as the difference between the optimum found by the optimization process and the true global optimum) may be limited by the computing time available. The ideal optimization algorithm would find a high quality solution, quickly, cheaply and use little memory. The computational effort is usually dominated by the evaluation of $f(\boldsymbol{x})$, and therefore the ideal algorithm is the often one which finds a solution using the fewest evaluations of $f(\boldsymbol{x})$.

In the n dimensional space described by $f(\boldsymbol{x})$ (often called the *feature space*), an extremum (maximum or minimum point) may be either *global* if it is the true largest or smallest value of the function, or

local if it is the largest or smallest value of the function within a finite space neighbouring the point. One is normally interested only in finding the global extremum, but this is, in general, an extremely difficult problem, and in many cases one has to settle for a close approximation to the true global minimum value. Assuming that we are dealing with minimization: the optimzation algorithm, starting at a given point will try to follow the value of the function $f(\boldsymbol{x})$ *downhill* towards the minimum. However, it is easy for the algorithm to become trapped in a local minimum – this is illustrated in Figure 6.1.

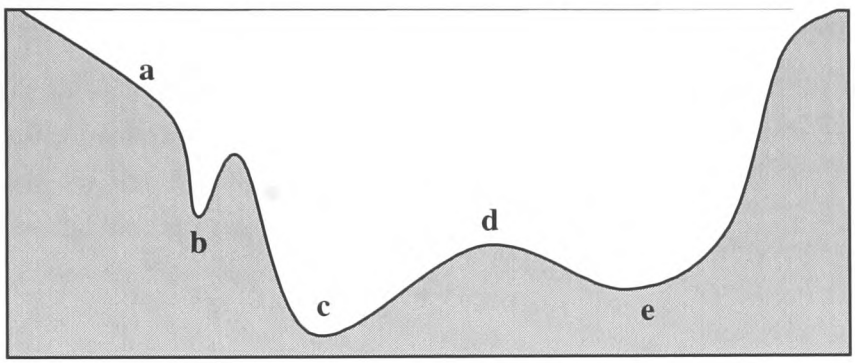


Figure 6.1 Global and local extrema of a one-dimensional function in a finite interval. Points b and e are local minima, point c is the global minimum. Starting at a, a simple gradient following minimization algorithm (hill climbing), would become trapped in the local minimum at b. Starting at d, the algorithm might proceed to either point c or e depending upon the precise value of the slope of the function in each direction.

Two standard heuristics to avoid the algorithm becoming trapped in local minima are widely used: (1) restart the optimization from a number of widely differing starting points and pick the lowest of these, and (2) perturb a local minimum by making a finite step away from it, restarting the optimization and seeing if the routine always returns to the same point. Any additional *a priori* information that can be utilized in the optimization may be of great benefit, increasing both the speed and accuracy of the process. The efficiency of optimization of models of natural systems such as vegetation canopies may be significantly improved by constraining the values of the parameters $\boldsymbol{x}$, to be within physically meaningful limits. For example, in canopy reflectance models, the sum of the leaf reflectance and transmittance terms must be less than 1. Knowledge of the likely vegetation type, or growth stage, can yield constraints on the maximum and minimum values for the leaf area index, which confine this parameter to a narrower range than would be possible without *a priori* knowledge.

Generally for optimizing models, such as models of vegetation reflectance, a function to be minimized, usually called the *merit function*, must be constructed from the output data of the model

and the measured data. For a model with *i* output variables, and for j sets of measured data, the merit function typically takes a form as follows:

$$f(\mathbf{x}) = \sum_i \left[\sum_j \left(R_{ij} - R_{ij}^{'} \right)^2 \right] \tag{6.1}$$

where: R_{ij} is the *i*th model output variable under the *j*th set of observing conditions

$R_{ij}^{'}$ is the *i*th measured parameter under the *j*th set of observing conditions

In the field of SVAT or BRDF modelling there has been a tendency to use black box optimization algorithms and, consequently few comparisons have been made of the success of different techniques. Moreover, since the goal of this research project was to develop a uniquely flexible SVAT modelling approach, whereby the model may be confronted with a variety of remotely sensed and non-remotely sensed data, in an automated fashion, optimization modules must in included in the code structure.

In order to compare the base performance of different optimization techniques a suite of algorithms was created, designed so that it may be incorporated into the linked model structure described in Chapter 5. The comparison was carried out through optimzation of the input parameters of the SAIL model by matching predicted and measured directional canopy reflectances. For complex, highly non-linear functions such as BRDF models, this is not an easy task. This forms one of the two main optimization paths for the linked model described in Chapter 5, and is therefore not only a tough test of the optimization algorithms, but also a good test of the validity of the linked modelling strategy.

In the linked model optimization step, the SAIL input parameters are derived from the other model components (SVAT, growth model etc.) in the system and/or other data (e.g. meteorological data). The SAIL model is then run in forward mode to predict canopy directional reflectances, and these are in turn compared with remotely sensed measurements. Any disagreement between the modelled and measured reflectances is then minimized by tuning the SAIL input parameters to minimize a merit function.

It is not at all straightforward to invert the SAIL model by rewriting the equations *in reverse*, therefore in order to optimize the model, repeated *guesses* must be made for the input parameters until the merit function is minimized. As described above, in order for the optimization to proceed at an acceptable rate a scheme must be devised to "home in" on the minimum as quickly as possible - rejecting spurious values for the model parameters.

Although the SAIL model has been successfully inverted by several authors, such studies have used a large number of reflectance measurements and because of the difficulties often encountered in locating suitable measured BRDF datasets, the measurements used in these studies are frequently *simulated measurements* obtained from forward running of the model, e.g. Goel and Thompson 1987. With in excess of fifty directional measurements several optimization techniques produce good minima for the merit function of the SAIL model, and hence an accurate estimation of the model parameters. However, such a large number of looks (i.e. reflectance measurements), whilst usually measusureable under controlled conditions during intensive field campaigns such as HAPEX-Sahel and BOREAS, are very unlikely to be available in practice from routine satellite soundings. Because of the limitations of engineering design and of the physics of satellite orbits, the number of looks for satellite instruments (i.e. view zenith directions) is probably be limited to a maximum of 10 all lying with a plane parallel to the direction of flight of the instrument. In contrast, building multi-spectral radiometers capable of making measurements in a large number of spectral bands is a relatively easy engineering exercise and therefore, whilst the number of looks may be limited, future sensor platforms have a large number of waveband channels available.

The merit function used in the optimization of the SAIL model was:

$$F = \sum_{i} (R_i - R_i')^2 \tag{6.2}$$

where i is a set of sun-observer geometries
R_i is the SAIL modelled reflectance at sun-observer geometry i
R'_i is the measured reflectance at sun-observer geometry i

The SAIL model assumes a horizontally homogeneous canopy (see Chapter 3), and is therefore best suited to the reflectance modelling of uniform canopies such as those of many arable crops, and is unsuited to the modelling of most clumped forest canopies. Unfortunately, the detailed measured datasets required for testing the optimization of the SAIL model are not widely available. In most studies measurements have been made over forest canopies; annual crops have rarely been investigated. As a consequence, the *"measured"* data in this study were simulated from forward running of the SAIL model.

In this procedure, a set of SAIL input parameters is created, and considered as the *true* parameter set. The model is then run in the forward directions to create a set of measured reflectance data, R'_i. Random or systematic noise may then be added to the measured reflectance data, through a random number generator, to simulate the noise that is always present in the signal from any instrument. The true parameter set is then *forgotten* and starting with an initial guess for the parameter set, R_i is

calculated from the model and the merit function F evaluated. The optimization algorithm is used to tune the input parameters to minimize the value of the merit function.

6.1. NUMERICAL ACCURACY

The numerical accuracy with which the result of an optimization can be found is clearly an important concern. This depends upon the computing hardware used, and the design of the software. However, most modern computers offer double precision arithmetic of high accuracy, in addition to single precision maths.

A first inspection might lead one to think that the greatest accuracy with which a minimum b of a function can be resolved is:

$$(1-\varepsilon)b < b < (1+\varepsilon)b \tag{6.3}$$

where ε is the numerical accuracy of the computer's floating point unit.

However, this is not the case. In general the shape of a function $f(x)$ near its minimum b will be given by the Taylor expansion (Press *et al.* 1992):

$$f(x) \approx f(b) + \frac{1}{2} f''(b)(x-b)^2 \tag{6.4}$$

On the right hand side of Equation 6.4 the second term will be negligible compared to the first term (and therefore $f(x) \approx f(b)$) when it is a factor ε smaller, i.e. when:

$$\varepsilon \frac{1}{2} f''(b)(x-b)^2 < f(b) \tag{6.5}$$

and therefore,

$$(x-b)^2 < \frac{2\varepsilon \cdot f(b)}{f''(b)} \tag{6.6}$$

which can be rewritten,

$$|x-b| < \sqrt{\varepsilon} \cdot |b| \cdot \sqrt{\frac{2|f(b)|}{b^2 f''(b)}} \tag{6.7}$$

When the above inequality holds, the point x will not yield a better minimum than the point b. The final square root term on the right hand side of equation 6.7 is of order unity for most functions, and therefore the bracketing interval $|x - b|$ must be greater than $|b|\sqrt{\varepsilon}$ for any improvement in the

minimum. In other words it is pointless attempting to bracket a minimum to within less than the square root of the machine precision (typically 3 x 10^{-8}, for single precision arithmetic on a Sun SPARC system).

6.2. DOWNHILL SIMPLEX

The downhill simplex technique was proposed by Nelder and Mead (1965). A simplex is a geometrical figure having (for N dimensions) N+1 vertices (Figure 6.2). So in two dimensions the simplex is a triangle, while in three dimensions it is a tetrahedron. A simplex is called non-degenerate if it encloses a finite volume of N-dimensional space. Each vertex of the simplex occupies a point in N-dimensional space representing one set of values for the parameters of the function to be optimized (the merit function). In the downhill simplex algorithm N+1 initial guesses are made for the parameters of the merit function; these guesses describe the initial vertices of the simplex. The value of the merit function is evaluated at each vertex. The algorithm then manipulates the simplex in a series of steps.

Figure 6.3 shows a flow chart of the downhill simplex algorithm, whilst the algorithm schematic described below.

Consider an N dimensional simplex with points $P_0, P_1, \ldots P_N$
Let y_i = value of merit function at P_i and $y_h = \max(y_i)$, $y_l = \min(y_i)$
Let C be the centroid of point with $i \neq h$, where h is the highest point.

For each step perform one of the above operations (reflection, contraction or expansion) and replace P_h (the hightest point) according to:

Reflection:

Let a new point be defined as: $P^* = (1+\alpha)\overline{P} - \alpha P_h \qquad (\alpha > 0)$

Where: α is the reflection coefficient chosen by trial and error.
If y^* is the value of the merit function at P^*, then if $y_l < y^* < y_h$ P_h is replaced by P^*.
If $y^* < y_l$ (i.e. reflection has produced a new minimum) then perform expansion.

Expansion:

Let a new point be defined as: $P^{**} = \gamma P^* + (1-\lambda)\overline{P} \qquad (\gamma > 1)$
Where γ is the expansion coefficient chosen by trial and error.

If $y^{**} < y_l$ replace P_h by P^{**} and restart

But if $y^{**} > y_l$ the expansion has failed, replace P_h by P^* and restart

Action	Description
Reflection	Move point of simplex with highest function value through opposite face of simplex. The volume of the simplex is preserved.
Expansion	If the reflection has produced a new minimum then the simplex is expanded in the new direction.
Contraction	At valley floors in the feature space describe by the merit function, the simplex is contacted toward the lowest point and tries to slide down the valley.
Multiple contraction	This is contraction in all directions and is employed when the simplex has to pass through a *hole*.

Table 6.1 The possible actions for each step of the Downhill Simplex algorithm

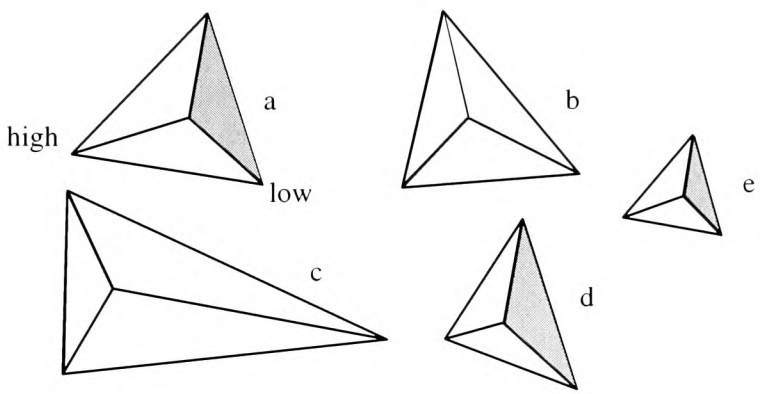


Figure 6.2 An illustration of the outcome of applying the four actions of Table 6.1 to a 3-dimensional simplex: a) the original simplex, b) reflection from the high point, c) reflection and expansion, d) 1-dimensional contraction away from the high point, e) multiple-dimensional contraction toward the low point.

The downhill simplex method is an elegant implementation of a gradient following technique. In general, it is good method for finding the minimum of a function in the absence of local minima. In terms of the number of function iterations, downhill simplex may not be the most efficient method, but it does find very close approximations to the minimum value for "smooth" functions (Spendley, Hext and Himsworth, 1962). As with many optimization methods it is necessary to make a number of restarts from different, random starting positions in order to avoid erroneous values caused by local minima.

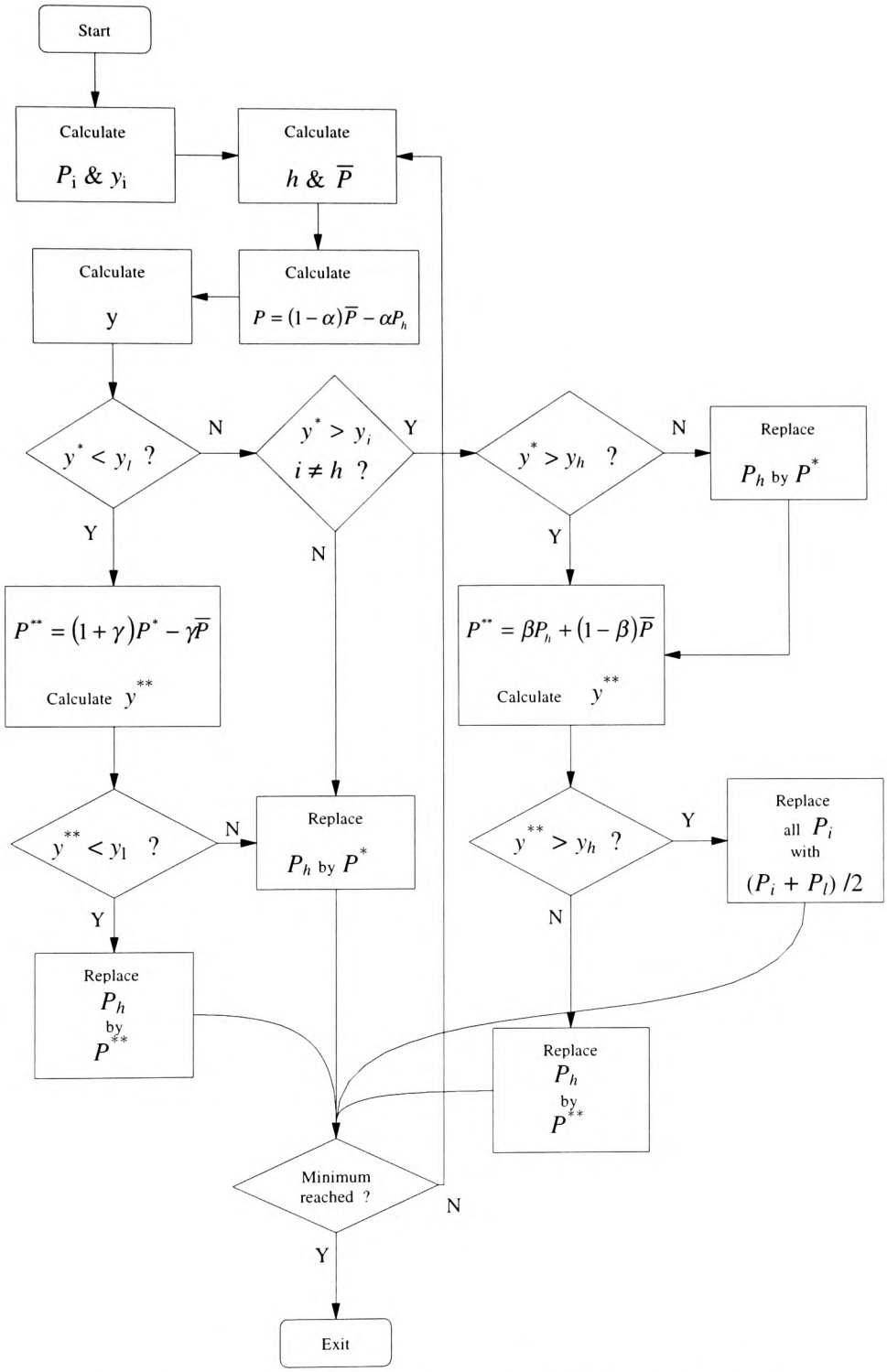


Figure 6.3 Flow chart of the Downhill Simplex optimization algorithm.

6.3. DIRECTION SET METHODS

6.3.1. Line Minimization

Consider the minimization of a function of one variable $f(x)$ (often called line minimization). Suppose that three abscissa are known, a, b and c, where $a < b < c$ (or $c < b < a$), such that $f(b) < f(a)$, and $f(b) < f(c)$. Then the function $f(x)$ must have a minimum in the interval (a, c) and the three points a, b and c are said to *bracket* the minimum. In order to find the minimum a new point, x, is chosen that it divides either the interval (a, b) or (b, c) and forms a new triplet of bracketing points. For example, suppose the point x was chosen to divide the interval (a, c). The new triplet of points is then (b, x, c). If $f(x) > f(b)$ then the new bracketing triplet would be (a, b, x), otherwise if $f(x) < f(b)$, the new triplet would be (b, x, c). In either case the middle point of the new bracketing triplet is the best minimum found so far. This process is then repeated until the distance between the two outer points of the triplet reaches some suitably small size.

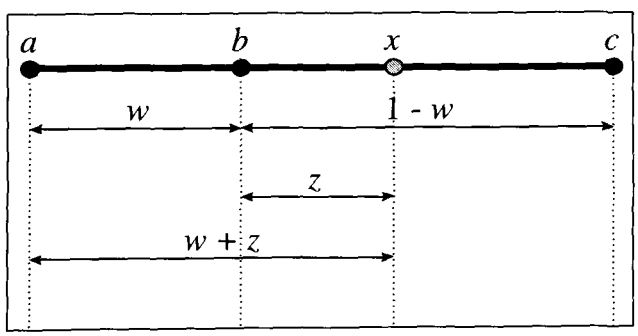
The optimum value for x can be found by considering the following.

Suppose b is a fraction w along the interval from a to c, i.e.

$$\frac{b-a}{c-a} = w \qquad \frac{c-b}{c-a} = 1-w \tag{6.8}$$

and suppose that the new trial point x is an additional fraction z beyond b, i.e.

$$\frac{x-b}{c-a} = z \tag{6.9}$$



The next bracketing triplet chosen will therefore have length $(w + z)$ if (a, x, b) is chosen, or $(1 - w)$ if (b, x, c) is chosen. To minimize the impact of the worst case scenario one chooses z to make these two segments have the same length:

$$z = 1 - 2w \tag{6.10}$$

This implies that x is symmetric to b in the original interval, i.e. that $| b - a | = | x - c |$. [Note also that x must be in the larger of the two segments, since $z < 0$ only if $w < ½$]. We now note that w is the trial

point from the previous iteration of the procedure! Therefore, w must have been optimally chosen in the same way as z above. Thus we assign x to be the same fraction of the way from b to c as b was from a to c. i.e.

$$\frac{z}{1-w} = w \tag{6.11}$$

Combining 6.10 and 6.11

$$w^2 - 3w + 1 = 0 \quad \text{which gives } w = 0.38197 \text{ and } \quad (1-w) = 0.61803 \tag{6.12}$$

These fractions are known as the Golden Ratio and have long been thought of in the world of art as having aesthetically pleasing qualities. The Golden Ratio is frequently used by artists when composing their works.

So for every iteration of the procedure, given a bracketing triplet of points (a, b, c), the next trial point is chosen to be a fractional distance 0.398197 into the larger of the two segments from the central point b. Each iteration of this procedure will bracket the minimum to an interval 0.61803 times the size of the previous one. This method if often called the Golden Search.

The Golden Search method is the most efficient way of hunting down the minima of awkward functions, however many functions are more co-operative. Brent (1973), introduced an algorithm (inverse parabolic interpolation) which is more efficient when the function being minimized is nearly parabolic in the region of its minimum. Inverse parabolic interpolation is best described with the aid of Figure 6.4. In Figure 6.4 the points 1,2 and 3 are an initial triplet of bracketing points. A parabola is *constructed* through these points and the merit function evaluated at the minimum of the constructed parabola, indicated by the arrow in the figure. This becomes a new point, point 4, which replaces the previous point with the largest function value, point 3. A new parabola is constructed through the new triplet 1,4,2. The minimum of the new parabola is point 5. It is easy to see that repeated iteration of this procedure will rapidly converge on the minimum of the merit function.

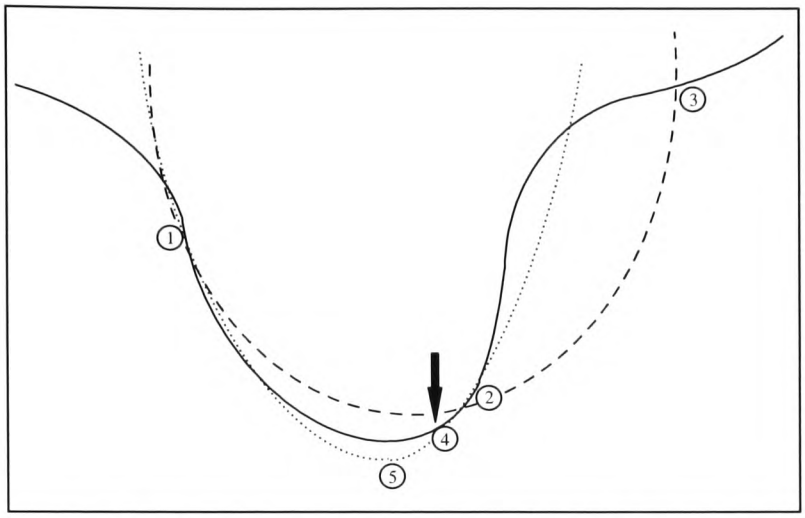


Figure 6.4 Brent's method of convergence to a minimum using parabolic interpolation. See text for explanation.

The equation giving the abscissa x which is the minimum or maximum of a parabola through three points a, b, c is:

$$x = b - \frac{1}{2} \cdot \frac{(b-a)^2 \cdot \big(f(b) - f(c)\big) - (b-c)^2 \cdot \big(f(b) - f(a)\big)}{(b-a) \cdot \big(f(b) - f(c)\big) - (b-c) \cdot \big(f(b) - f(a)\big)} \tag{6.13}$$

Brent (1973) describes an algorithm which uses the Golden Search when the behaviour of the merit function is awkward, but which switches over to parabolic interpolation whenever possible. This algorithm is very complex but the salient points are illustrated in Figure 6.5.

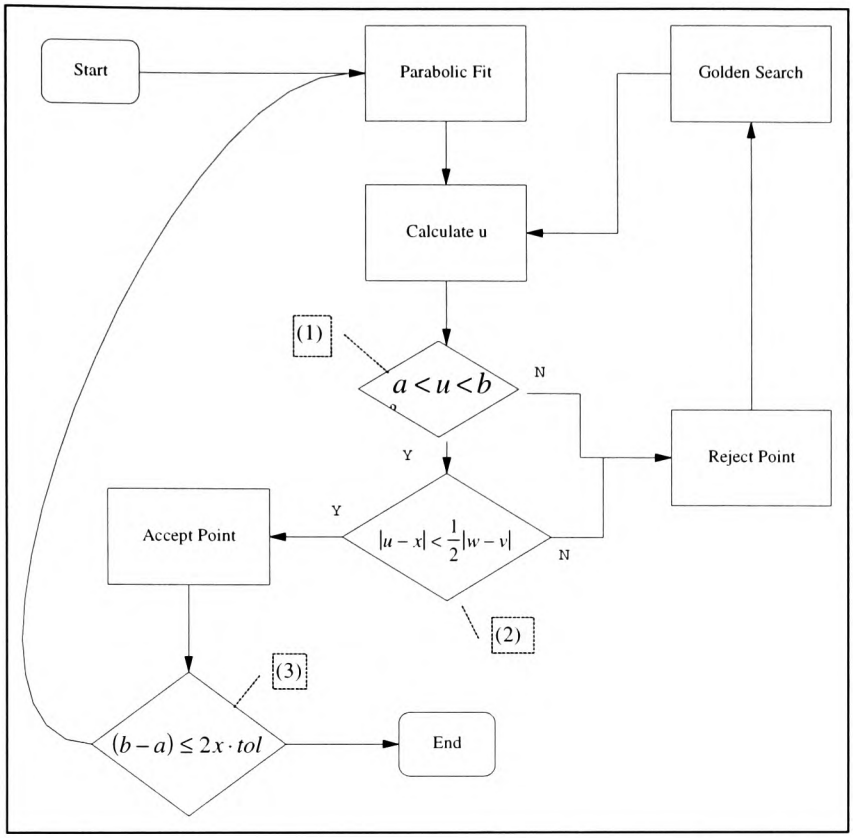


Figure 6.5 A flow diagram of Brent's efficient algorithm for finding the minimum of a function of one variable.

Define points: *a, b* - minimum bracketing points
x - best value for minimum found so far
w - second best value for minimum
v - previous value for w
u - most recent function evaluation

1) Point rejected if it is not within the interval *a, b*.
2) Point rejected if the distance of the new point *u* from the best point *x* is not less than half the distance of the step before last. The choice of the step before last is chosen to avoid punishing the algorithm for a single bad step.
3) Termination criterion after Brent (1973).

6.3.2. Multi-Dimensional Minimization

Consider extending the ideas of line minimization to a function of *N* variables, $f(\boldsymbol{x})$. Starting at a point $\boldsymbol{P}$ the function can be minimized in any given vector direction $\boldsymbol{n}$ using the line minimization algorithm. Multidimensional minimization could then be achieved by carrying out successive line minimizations in suitably chosen different directions. A simple method of choosing the new directions would be to take the unit vectors of $\boldsymbol{x}$, $\boldsymbol{u}_1$, $\boldsymbol{u}_2$,..., $\boldsymbol{u}_N$ as the set of directions, find the minimum in the first direction, $\boldsymbol{u}_1$, and *from there* move along the second direction $\boldsymbol{u}_2$ to its minimum and so on. However, whilst this method may work well for many functions, it works very badly for others, for example a function where a long narrow valley exits at some angle to the unit axes. In such

a case only very small steps can be made in the unit vector directions before hitting the walls of the valley, forcing may changes of direction and making progress toward the minimum very slow.

A better set of directions is required which either enable good progress along narrow valleys or else provide some "non-interfering" directions (called conjugate directions) which have the property that minimization along one direction is not spoiled by subsequent minimization along another.

A function $f(\boldsymbol{x})$ can be expanded into its Taylor series:

$$f(\boldsymbol{x}) = f(\boldsymbol{P}_0) + \sum_i \frac{\partial f}{\partial x_i} x_i + \frac{1}{2!} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j} x_i x_j + \cdots \tag{6.14}$$

$$\approx c - \boldsymbol{b} \cdot \boldsymbol{x} + \frac{1}{2!} \boldsymbol{x} \cdot \boldsymbol{A} \cdot \boldsymbol{x} \tag{6.15}$$

Where: $c \equiv f(\boldsymbol{P}_0) \qquad \boldsymbol{b} \equiv -\nabla f\big|_{P_0} \qquad \boldsymbol{A} \equiv \frac{\partial^2 f}{\partial x_i \partial x_j}\Big|_{P_0}$

From equation 6.15 the gradient of $f(\mathbf{x})$ is :

$$\nabla f = \boldsymbol{A} \cdot \boldsymbol{x} - \boldsymbol{b} \tag{6.16}$$

Note that the gradient is zero and therefore at either a maximum or minimum point when $\boldsymbol{A} \cdot \boldsymbol{x} = \boldsymbol{b}$ .The change in gradient along some vector direction $\delta\mathbf{x}$ is:

$$\delta(\nabla f) = \boldsymbol{A} \cdot (\delta \boldsymbol{x}) \tag{6.17}$$

Consider an algorithm which has successfully found a minimum of $f(\boldsymbol{x})$ in the direction $\boldsymbol{u}$. If the proposed new direction along which to minimize $\boldsymbol{v}$ is not to spoil the minimization already achieved in the direction $\boldsymbol{u}$, the gradient of $f(\boldsymbol{x})$ must remain parallel to $\boldsymbol{u}$, i.e.

$$\delta(\nabla f) \cdot \boldsymbol{u} = 0 = \boldsymbol{u} \cdot \boldsymbol{A} \cdot \boldsymbol{v} \tag{6.18}$$

When equation 6.18 holds for two vectors **u** and **v**, then these directions are said to be conjugate. Successive line minimizations along conjugate directions do not require repetition and one pass on N minimizations will find the exact minimum of quadratic functions of the form of equation 6.14. For functions that are not exactly quadratic the exact minimum will not be found after one pass but repeated passes of the algorithm will converge quadratically toward the minimum.[4]

[4] Steepest descent methods exhibit only linear convergence on a quadratic function and so are not quadratically convergent.

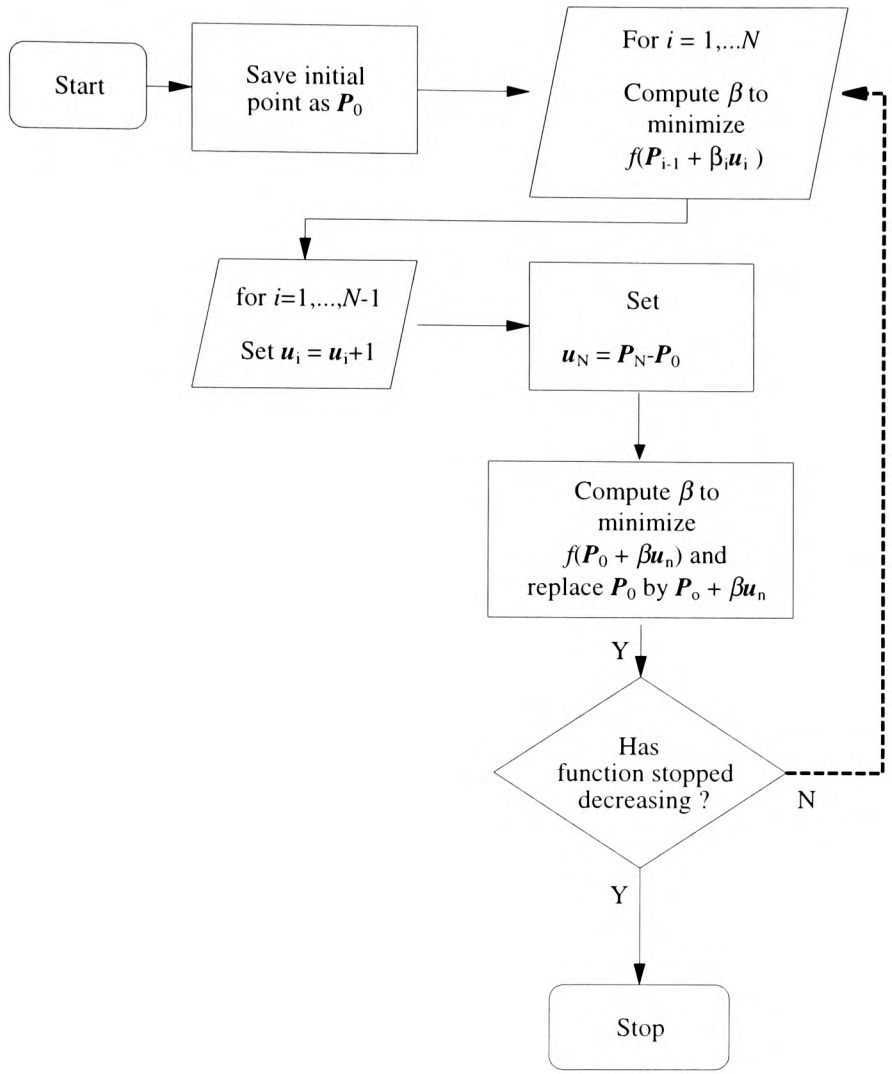


Figure 6.6 A flow diagram showing Powell's direction set optimization algorithm (Powell 1964).

Powell (1964) proposed a method that produces N mutually conjugate directions, the flow diagram of his algorithm being shown in Figure 6.6.

It can be shown that N iterations of the above algorithm, and therefore $N(N + 1)$ line minimizations in total, will exactly minimize a function which can be expressed in a quadratic form like Equation 6.15 (Brent 1973). However, there is a problem with this algorithm, which is that for any iteration the value of β_1 might be zero. This results in the directions $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n$ becoming lineally dependent and only a subsection of the N-dimensional parameter space being searched. Even though it is unlikely that β_1 would vanish exactly, in practice the directions $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n$ often become nearly linearly dependent. A number of suggestions have been made as to possible solutions to this problem. Powell

(1964) suggested the the new direction ($\boldsymbol{P}_n - \boldsymbol{P}_0$) should only be used if this does not increase the value of $\left|\det(v_1,\ldots,v_n)\right|$, where

$$v_i = \left(\boldsymbol{u}_i^T A_u i\right)^{-1/2} \boldsymbol{u}_i \tag{6.19}$$

This modification is successful, but loses the property of quadratic convergence. The simplest way to avoid linear dependence of the search directions and maintain quadratic convergence for $\beta > 0$ is to reset the search directions to the columns of the identity matrix every n iterations. Unfortunately, such *restarts* may slow convergence because information obtained about the function is thrown away. Brent (1973) suggested a method that avoids discarding useful information. Instead or resetting the search directions to the identity matrix they can equally well be reset to the columns of any orthogonal matrix $\boldsymbol{Q}$. To avoid discarding useful information, $\boldsymbol{Q}$ can be chosen so that the search directions remain conjugate if $f()$ is quadratic. Brent's procedure for calculating the principal vectors $\boldsymbol{q}_1,\ldots,\boldsymbol{q}_n$ which form the new search direction uses a single value decomposition approach, and although elegant, is rather complex and therefore will not be stated here – refer to Brent (1973) pp 129-131.

For the purposes of this project a simpler technique, proposed by Acton (1990), has been adopted. This techniques abandons quadratic convergence and instead modifies the Powell procedure to attempt to find several directions which are parallel to narrow valleys in the parameter space, rather than finding N mutually conjugate directions along which to minimize.

Consider a situation where the topography of the parameter space twists repeatedly from side to side. In such a case it is not easy to find N mutually conjugate directions, nor is the situation suited to parabolic extrapolation. Acton's algorithm retains ($\boldsymbol{P}_N$ - $\boldsymbol{P}_0$) as the new direction; ($\boldsymbol{P}_N$ - $\boldsymbol{P}_0$) is the average direction moved after investigating N possible alternatives and is likely to be reasonable well oriented with valleys in the parameter space. However, instead of discarding $\boldsymbol{u}_N$, the direction discarded is the one for which the merit function exhibited the largest decrease on the previous iteration $\boldsymbol{u}_B$. This may seem to be throwing away all that was gained on the previous iteration, but $\boldsymbol{u}_B$ is a major component of ($\boldsymbol{P}_N$ - $\boldsymbol{P}_0$) and therefore discarding it minimizes the chances of creating linear dependence between directions.

However, the old direction is kept and ($\boldsymbol{P}_N$ - $\boldsymbol{P}_0$) is not used as the new direction in the following two cases: Define $f_x \equiv f\left(\boldsymbol{P}_x\right)$ and $f_E = f\left(2\boldsymbol{P}_N - \boldsymbol{P}_0\right)$ (6.20)

1) $f_E \geq f_0$ - the extrapolated direction f_E has been exhausted as a source for better minima.

2) $2\left(f - 2f_N + f_E\right)\left[\left(f_0 - f_N - \Delta f\right)\right]^2 \geq \left[f_0 - f_E\right]^2 \Delta f$ - either the decrease along $(\boldsymbol{P}_N - \boldsymbol{P}_0)$ was not primarily because of the decrease in any single direction, or, the second derivative along $(\boldsymbol{P}_N - \boldsymbol{P}_0)$ is significantly large and a close minimum has already been achieved.

6.3.3. Summary

Direction set methods, such as the one described here, have proved to be efficient at finding minima in a wide range of optimization problems (Brent 1973). However, complex multi-dimensional functions with twisty valleys and local minima can confuse the algorithm and make the selection of new search directions difficult. This may result in only a sub-volume of the feature space being searched, and the global optimum escaping detection (Acton 1970). Restarting the algorithm for different random starting position may reduce this problem, but at the expense of efficiency. This technique remains, however, one of the most powerful non-stochastic optimization methods.

6.4. SIMULATED ANNEALING

This simple and elegant technique has received increasing attention in recent years because of its ability to avoid becoming trapped in local minima. The idea for the technique came from the modelling of thermodynamic systems by Metropolis *et al.* (Metropolis *et al.* 1953). Kirkpatrick *et al.* (Krikpatrick 1984; Kirkpatrick *et al.* 1983) developed Metropolis' model into an algorithm for the minimization of functions.

The simulated annealing (SA) technique is analogous to the process of cooling and annealing in metals. Consider some molten metal that is cooling.

At high temperatures the molecules within a liquid metal are in rapid random motion relative to one another. As the liquid cools, the thermal motion of the molecules slows and the energy of the system is reduced. If the cooling is sufficiently slow the molecules are able to align themselves into ordered crystal lattices, which are the minimum energy state for the metal. However if the liquid metal is cooled too rapidly the molecules do not have time to order themselves into crystals and instead form an amorphous structure with a somewhat higher energy state. Thus only if the cooling is sufficiently slow is the lowest energy state achieved.

There is an obvious analogy with the optimization of functions; most traditional optimization methods attempt to reach a minimum value rapidly by taking the fastest possible route downhill. This often leads to local minima analogous to the amorphous state of metals. A slower *cooling* is required in order to obtain a global minimum.

The probability of a system at a given temperature being in a given energy state is given by the Boltzmann distribution:

$$\text{Prob}(E) \approx \exp\left(\frac{-E}{kT}\right) \tag{6.21}$$

where k is the Boltzmann constant.
T is the temperature.
E is the energy.

Therefore, even at low temperature, there is a small probability of the system being in a relatively high energy state. Thus to get out of local (energy) minima, natural systems occasionally take *uphill* steps as well as downhill; the lower the temperature the lower the probability of moving uphill.

Metropolis *et al.* (1953) developed a simulation of such a thermodynamic system. In their model the system changed from energy state E_1 to E_2 with a probability given by:

$$\exp\left[\frac{-(E_2 - E_1)}{kT}\right] \tag{6.22}$$

Note that for $E_2 < E_1$ the probability is greater than 1 and so this change of state always occurs (a downhill step is always taken). When $E_2 > E_1$ an uphill step is taken with a probability given by Equation 6.22, so that a change of state from one of low energy to one with a higher energy sometimes occurs. Thus, such a system is able to leap out of local minima, avoiding becoming trapped. The probability of uphill steps being accepted is higher at higher temperatures. As the temperature of the system is lowered uphill steps become less and less likely until when $T = 0$ the system reduces to a simple gradient search method.

In order to apply these ideas to the minimization of a function of N variables we need he following:

- an "energy" function, E, which is to be minimized - this is the *merit* (or *cost*) function,
- a way of generating random changes in the system configuration,
- an initial state for the system, and
- an initial value for the temperature parameter T, and an "annealing schedule" which determines how to reduce the temperature.

6.4.1. Implementation of the simulated annealing algorithm

Early applications of the simulated annealing technique were in a variety of combinatorial optimization problems such as integrated circuit design, and SA was the first technique successfully to solve the famous travelling salesman problem. Combinatorial problems involve finding optimum

permutation of a discrete set of parameters. Whilst such problems are numerous, continuous optimizations problems, where the optimum values of a set a continuously varying parameters is sought, also frequently arise, and it is continuous optimization that we are concerned with in this study.

The main complication introduced when considering continuous as opposed to discrete optimization problems is that the choice of the random changes in the system, or steps, becomes more subtle. Steps that are too small will be very inefficient at exploring the feature space, steps that are too large will usually be rejected, hindering progress toward the global minimum. Therefore, one of the key factors contributing to the efficiency of the algorithm is how to generate the random changes in the system. A generator is required which (nearly) always takes a downhill step if one is available. Another problem is how slowly to reduce T, too quickly and the system will reach an amorphous state which is not the true minimum, too slowly and the time taken for the procedure will become prohibitive. Choosing the initial value for the temperature of the system T_0 requires a little care. If T_0 is too low the system will not start in a molten state, and the whole of the feature space may not be searched, while if T_0 is too high the number of iterations may become excessive or the temperature may have to be lowered too quickly.

In this study, two different implementations of the basic simulated annealing algorithm were developed. In the first method, a self-regulating step distribution algorithm is employed. The second method is a modification of the technique proposed by Press *et al.* (1992) in which the downhill simplex technique forms the basis of the search.

6.4.2. Self-regulating step distribution (SA1)

Consider a merit function expressed as $E(\boldsymbol{x}) = E(x_1, x_2, ..., x_n)$.

Proceeding from the starting point the simulated annealing algorithm makes random steps $\Delta\mathbf{x}$. For each step

$$\Delta E = E(\boldsymbol{x} + \Delta\boldsymbol{x}) - E(\boldsymbol{x}) \tag{6.23}$$

The step is accepted with a probability p defined by:

$$p = \exp\left[\frac{-\Delta E}{T}\right] \tag{6.24}$$

and if $\Delta E < 0$ the step is always accepted.

The series of steps generates a random walk that can be expressed as:

$$P(\mathbf{x}) = \frac{1}{\int d^n x \exp\left[\frac{-E(\mathbf{x})}{T}\right]} \cdot \exp\left[\frac{-E(\mathbf{x})}{T}\right] \tag{6.25}$$

where: $P(\mathbf{x})d^n x$ is the probability that the walk will be in the volume $d^n x$ on a given step.

The random steps are chosen according to:

$$\Delta \boldsymbol{x} = \boldsymbol{Q} \cdot \boldsymbol{u} \tag{6.26}$$

where: $\boldsymbol{u} = (u_1, u_2, ..., u_n)$ a vector of random numbers with zero mean and unit variance and $\boldsymbol{Q}$ is the step distribution matrix.

The step distribution given by $\boldsymbol{Q}$ can be characterised in terms of a covariance matrix **s**. It can be shown that (Vanderbilt and Louie 1984):

$$\boldsymbol{s} = \boldsymbol{Q} \cdot \boldsymbol{Q}^T \tag{6.27}$$

The efficiency of the annealing algorithm depends upon the schedule for reducing the temperature T, and distribution of the random steps, given by the covariance matrix s. A random walk with any desired covariance matrix **s** can be generated by solving Equation 6.27 for $\boldsymbol{Q}$. Because optimization is such a computationally intensive process, the fastest possible (i.e. computationally efficient) solution is sought at each stage of the procedure. For equations such as 6.27 this is the Cholesky decomposition; however, it cannot be used in this instance since it requires that s be both symmetric and positive definite. A matrix **s** is positive definite when:

$$\boldsymbol{v} \cdot \boldsymbol{s} \cdot \boldsymbol{v} > 0 \quad \text{for all vectors } \boldsymbol{v} \tag{6.28}$$

Unfortunately, Equation 6.27 does not always hold in our case, therefore the slower (by a factor of two) LU decomposition technique must be used.

Choosing the covariance matrix

The optimal covariance matrix is one where for M steps the random walk explores most of the phase space $\Omega(T)$ available to it at temperature T. Both the size and the shape of the steps taken are important. If the steps are too small all the steps will be accepted, if the steps are too large, all will be rejected. In both cases little information is gained per iteration. The optimum step size is one for which about half the steps are accepted. However, the region of parameter space explored (governed by s) is also important. If s and $E(t)$ are not well aligned a large number of iterations will be wasted searching in useless directions. Vanderbilt and Louie (1984) proposed a scheme that uses the shape of the random walk itself as a measure of the topography of $E(t)$ as follows.

After the lth set of M steps, the first and second moments of the segment of the random walk are calculated from:

$$A_i^{(l)} = \frac{1}{M} \sum_{m=1}^{M} x_i^{(m;l)} \tag{6.29}$$

$$S_{ij}^{(l)} = \frac{1}{M} \Sigma \left[x_i^{(m;l)} - A_i^l \right] \left[x_j^{(m;l)} - A_j^l \right] \tag{6.30}$$

Where: $x^{(m;l)}$ is the value of x on the m^{th} step of the i^{th} set

The matrix $\boldsymbol{S}$ describes the shape of a segment of the walk ($\boldsymbol{s}$ describes the probability distribution of the individual trial steps.) For a walk where no steps are rejected (free walk) we have:

$$\left\langle \mathbf{S}_{free}^l \right\rangle = \beta M \mathbf{s}^l \tag{6.31}$$

Where: $\left\langle \mathbf{S}_{free}^l \right\rangle$ indicates the average of $\boldsymbol{S}$ over the nM random variables $\Delta x_i^{(m;l)}$

$\beta = 0.11$ from calculation (see Vanderbilt and Louie 1984).

s for the (l+1) set is then calculated from:

$$\mathbf{s}^{(l+1)} = \frac{\chi_s}{\beta M} \mathbf{S}^{(l)} \tag{6.32}$$

and

$$\beta = 0.11 \tag{6.33}$$

Where: χ_s is the "growth factor" (chosen > 1) so that each set covers $\chi_s^{1/2}$ times as much space in each direction as the previous set.

Thus, the step size distribution is recalculated after each set of M steps from equations 6.29, 6.30 and 6.32 and the new steps are calculated from equations 6.27 and 6.26.

The algorithm works as follows: if s is initially too small, all the steps are small and all are accepted. In this case, equations 6.31 and 6.32 apply and so the step size grows by $\chi_s^{1/2}$. Eventually the step size grows until the limits imposed by $E(T)$ are reached; at this point a large number of steps start to become rejected. Equation 6.31 no longer holds and $\boldsymbol{S}$ and therefore $\boldsymbol{s}$ begin to reflect the shape of $E(T)$. As T is reduced, $\boldsymbol{s}$ is maintained at an appropriate size relative to $E(T)$. Thus if a large number

of steps are rejected in a given direction, the random walk is constrained in that direction but will have large excursions along other directions.

M must be chosen to be sufficiently large so that there are reasonable statistics in Equation 6.30. Setting $M \approx 15n$ seems to work well for problems with less than about nine dimensions. Note that there is a further problem with this algorithm if the number of dimensions, n, is large. Since S must have n eigenvalues, if n is very large it is likely that at least one of the eigen values will be very small. The exploration of the direction characterised by the corresponding eigenvector will consequently be constrained.

Vanderbilt and Louie (1984) suggest possible solutions for this problem, however, the maximum number of dimensions encountered in all the simulations was less than ten, and no corrections to the algorithm were found to be necessary.

The termination criterion used for this algorithm was:

$$\frac{\left\langle E^{(m)} \right\rangle - E_{\min}^{(m)}}{\left\langle E^{(m)} \right\rangle} < \eta \tag{6.34}$$

where $\left\langle E^{(m)} \right\rangle$ is the average of E over M steps of the random walk and

$E_{min}^{(m)}$ is the minimum value of E over the M steps.

6.4.3. SA with downhill simplex search (SA2)

In this technique, the single vector x representing the search point in the feature space is replace by a simplex with $N+1$ vertices. The operations on the simplex are the same as those described in Section 6.2, namely reflection expansion and contraction. The implementation of the SA algorithm is rather subtle and follows Press *et al.* (1992). A positive, logarithmically distributed random number, proportional to the temperature is added to the value of the merit function at each vertex of the simplex. A similar random number is subtracted from the value of the merit function at every new trial point selected. This algorithm always takes a downhill step, but also sometimes accepts an uphill one, thus following the concept of the Metropolis procedure (Equation 6.22).

Choosing an annealing schedule

The choice of an annealing schedule to reduce the temperature T can have a significant effect on the performance of any SA algorithm. Two different schedules were adopted in this study, the "alpha" schedule given by:

$$T = T_0\left(1 - \frac{k}{K}\right)^{\alpha} \tag{6.35a}$$

Where: T_0 is the initial temperature

K is the maximum number of function evaluations allowed

k is the cumulative total of evaluations

α is a constant

and the "logarithmic" schedule, based on the following strategy:

Set the total number of trials allowed, K

Set $K_T = K^P$

For i = 1 to K_T let

$$T_i = W^{\left(\frac{N}{C} - i\right)} \text{ and do } \frac{K}{K_T} \text{ function evaluations at this temperature.} \tag{6.35b}$$

Where: $0.0 < P < 1.0$ with 0 implying a random search at a given temperature and 1.0 implying a single trial at each temperature.

$1.0 < W < 10.0$ with large values implying a wide range of temperatures, 1.0 fixes all temperatures to be the same

$0.1 < C < 10.0$ with large values biasing the temperature range toward lower, colder values, 2.0 biasing the temperatures equally between hot and cold.

Unless computing resources are unlimited, it is usual to set the maximum number of iterations of the algorithm, K. The maximum number of function evaluations at each temperature, f_{lev}, is usually also set which determines k in Equation 6.35a. The parameter α determines how the iterations are spread between high and low temperatures. Larger values for α imply using more iterations at lower temperatures. For $\alpha = 1$, the temperature is reduced linearly: α typically takes on values between 0.5 and 4.0.

The setting of the initial temperature T_0 can be problematic. Kirkpatrick *et al.* (1983) suggest that there is often a crucial temperature around which the system begins to freeze, and where slow careful annealing is beneficial in finding the true minimum. They define a quantity ($C(T)$) analogous to specific heat:

$$C(T) = \frac{d}{dt}\langle E(T)\rangle = \frac{1}{T^2}\left[\langle E^2\rangle - \langle E\rangle^2\right] \tag{6.35c}$$

A peak in the value of the specific heat indicates that the system is crystallizing or becoming ordered. For complete annealing it is necessary for the system to be molten initially: this is achieved by setting the initial temperature T_0 to a "sufficiently high" value. One strategy for determining how high is "sufficiently high" is to increase the temperature from a very low value until the peak in the value of the specific heat has been passed, thus indicating a change of state from the solid to the liquid phase. This temperature is then used for T_0. The rate of temperature reduction can be varied during the optimization in an attempt to improve performance. The peak in the value of the specific heat indicates that the system is becoming ordered. Slower annealing at this temperature is sometimes useful to ensure that the annealing optimization does not become trapped in a local minimum.

6.4.4. Summary

Simulated annealing has been receiving increasing attention in recent years, because of its unique features. Applications have illustrated its ability to avoid becoming trapped in local minima, but results indicate that careful choice of the controlling parameters is required to obtain the best from this technique (Szu and Hartley 1987, Connolly 1992). Like other stochastic algorithms such as genetic algorithms (see section 6.5), the major disadvantage of SA is the heavy computational burden it involves.

6.5. GENETIC ALGORITHMS

Genetic algorithms (GAs) are optimization techniques that attempt to mimic some of the processes of evolution and natural selection (Holland 1987). Evidence for the success of natural selection is all around us, and may be observed in the exquisite adaptation of many species to their environment. In nature, each species adapts to a complicated and changing environment in order to maximise the likelihood of its survival. The knowledge that each species gains is encoded in its genes. Over time, changes to the genes give rise to individuals that are more likely to survive, and so have a greater chance of passing their improved characteristics on to future generations. Of course, not all changes will be beneficial but those which are not tend to die out. This idea forms the basis of the Darwinian concept of "Survival of the Fittest".

In genetic algorithms, the genes are represented by the parameters of the function to be optimized and the merit function represents the measure of the fitness of a set of parameters. For each generation

(iteration of the GA) members of a population of possible values of the parameters set are selected and combined in some, as yet unspecified, way. After each successive generation a closer and closer approximation to the optimum parameter set for the merit function should be given by the fitness chromosome in the population.

Genetic algorithms differ from *normal* optimization and search techniques in the following three ways:

1. GAs work with a coding of the parameter set, not the parameters themselves.
2. GAs use a relatively large population of points rather than a single point or connected group of points.
3. GAs use probabilistic non-deterministic transition rules.

Genetic algorithms require the parameter set being optimized to be encoded as a finite-length string, the elements of which can take on some value from a specified finite range or alphabet. The individual elements of the string are known are genes, and the string of genes, which represents a solution, is known as a chromosome. Individual genes take on one of a number of values, called alleles, drawn from a fixed alphabet. The position of a gene along the chromosome is known as its locus.

The standard genetic algorithm proceeds as follows: an initial population of individuals is generated at random or heuristically. At each iteration of the algorithm, known as a generation, the individuals in the current population are decoded and evaluated according to some predefined quality criterion, known as the fitness function. Individuals are selected for reproduction according to their fitness values. Many selection procedures may be used, but frequently individuals are selected with a probability proportional to their relative fitness. This ensures that the expected number of times an individual is chosen is approximately proportional to its relative performance in the population. Thus, high-fitness individuals are more likely to reproduce than low-fitness ones.

Selection alone does not introduce any new individuals into the population, i.e., it cannot find new points in the search space. These are generated by the crossover and mutation *operators*. Crossover is performed with probability p_c between two or more selected individuals, called parents, by exchanging parts of their genomes to form two new individuals, called offspring. This operator tends to enable the evolutionary process to move toward promising regions of the search space. The mutation operator is introduced to prevent premature convergence to local optima by randomly sampling new points in the search space. It is carried out by altering at random parts of the chromosome with some (small) probability p_m. Genetic algorithms are stochastic iterative processes that are not guaranteed to converge; the termination condition may be specified as some fixed, maximal number of generations or as the attainment of an acceptable fitness level.

The basic genetic algorithm proceeds as follows:

- an initial population of chromosomes is constructed at random,
- at each generation, the fitness of each chromosome in the population is measured, by evaluating a merit function,
- a selection routine picks a group of the fittest chromosomes which are used to produce children by sexual reproduction,
- offspring inherit characteristics from both parents, but random mutations may be introduced during reproduction, and
- after many generations of selection for the fitter chromosomes, the result is, hopefully, a population that is substantially fitter than the original.

Chromosomal Representation

Each chromosome represents a solution to the problem and is composed of a string of genes. The binary alphabet {0,1} is often used to represent these genes but sometimes, depending on the application, integers or real numbers are used. In fact, almost any representation can be used that enables a solution to be encoded as a finite length string.

Initial Population

Once a suitable representation has been decided upon for the chromosomes, it is necessary to create an initial population to serve as the starting point for the genetic algorithm. This initial population can be created randomly or by using specialised, problem specific, information. From empirical studies, over a wide range of function optimization problems, a population size n of between 30 and 100 is usually recommended (Goldberg and Deb 1991).

Fitness Evaluation

Fitness evaluation involves defining an objective or fitness function against which each chromosome is tested for suitability for the environment under consideration. As the algorithm proceeds the individual fitness of the "best" chromosome should increase as well as the total fitness of the population as a whole.

Reproduction

This is the process whereby individual chromosomes are copied according to their merit function values. Copying chromosomes according to their fitness means that chromosomes with a higher fitness value will have a higher probability of contributing one or more offspring in the next generation. In the simplest GA the new generation entirely replaces the old one, but various modifications to this behaviour are possible.

Selection

The selection of chromosomes for reproduction may be made in a variety of ways, one of the simplest algorithms is the so-called *Biased Roulette Wheel.* In this selection algorithm, each chromosome occupies space on a roulette wheel in proportion to its fitness. Selection is then achieved by spinning the roulette wheel n times. Each selected chromosome is then cloned, a process in which an exact copy of the chromosome is placed in a mating pool to undergo crossover and mutation operations.

Crossover

Once a pair of reproduced chromosomes has been selected, crossover can take place to produce offspring. Based on a specified crossover probability, members of the newly reproduced chromosomes in the mating pool are mated at random. Each pair of chromosomes then undergoes crossover. Consider two chromosomes A1 and A_2 of length l (Figure 6.7). During crossover an integer position (the crossover locus) k along the chromosome is selected at random between 1 and $l-1$. Two new chromosomes are created by swapping all the genes between positions $k+1$ and l inclusive.

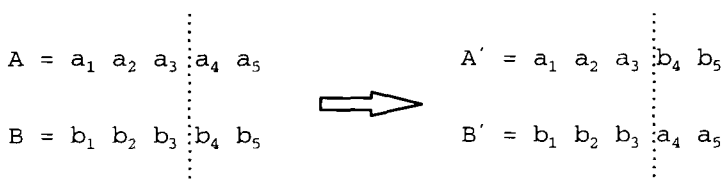


Figure 6.7 Crossover at position $k = 4$ between two chromosomes of length $l = 5$. A andB are the chromosomes before, A' and B' and the chromosomes after crossover.

A crossover probability of 1.0 indicates that all the selected chromosomes are used in reproduction i.e. there are no survivors. However, better results are usually achieved using a crossover probability of between 0.65 and 0.85 (Booker 1987), which implies that the probability of a selected chromosome surviving to the next generation unchanged (apart from any changes arising from mutation) ranges from 0.35 to 0.15. Crossover can be made to occur at more than one point along the chromosome; two point crossover crosses the parents at two randomly chosen points.

Mutation

If the crossover operator alone is used to produce offspring, one potential problem that may arise is that if all the chromosomes in the initial population have the same value at a particular position then all future offspring will have this same value at this position. For example, if all the chromosomes have a 0 in position two then all future offspring will have a 0 at position two. To combat this undesirable situation a mutation operator is used. This attempts to introduce some random alteration of the genes, e.g. 0 becomes 1 and *vice versa.* Typically, this occurs infrequently so mutation is of the order of about one bit changed in a thousand tested. Each bit in each chromosome is checked for

possible mutation by generating a random number between zero and one and if this number is less than or equal to the given mutation probability, e.g. 0.001, then the bit value is changed.

Various subtle refinements can be made to the basic GA in an attempt to improve performance (Goldberg 1987, Keane, 1995). For example, elitism can be introduced where the fittest member, or group of members, from each generation is carried over into the next generation. Similarly, generators can be made to overlap by specifying a percentage of the population that should be replaced in each generation. In contrast to the simple GA described above, newly generated offspring are added to the population and an equal number of individuals are destroyed, thus maintaining the size of the population. The new offspring are typically made to replace either their parents, a random member, or the worst member of the previous generation. Of course, there is no reason why offspring should be the result of reproduction between only two parents, nor is there any restriction on the number of children that result from a mating. One refinement that may be introduced in order to give the crossover performance an added effectiveness is to perform a quick local descent with each offspring after the crossover and mutation (Goldberg and Deb 1991). Quick local descent modifies the current offspring to a neighbouring offspring with lower cost. In contrast to local steepest descent, it interrupts immediately after a neighbour that has a lower cost is found. This does not guarantee to find the best neighbour, but it is significant faster than the local steepest descent and still an important improvement compared with the original offspring.

6.5.1. A simple example

Consider an attempt to find the maximum possible value of a "black box" signal generator. The value of the signal, $f(x) = x^2$ is a function of the position of five binary switches, as illustrated in Figure 6.8.

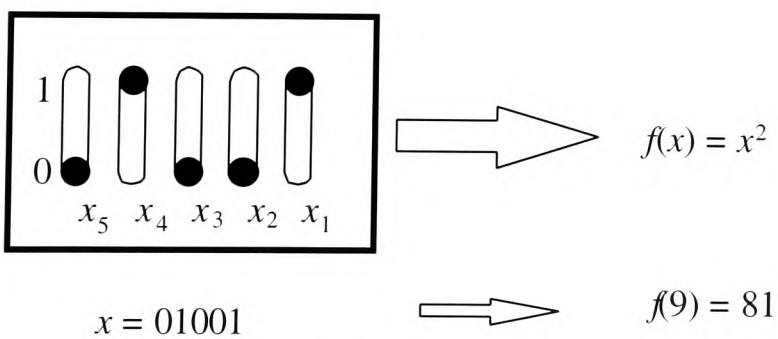


Figure 6.8 A *black box* optimization problem.

This problem clearly lends itself to a binary string encoding, where 0 represents a closed switch and 1 an open switch. An initial population for chromosomes is generated by randomly choosing the value (0 or 1) for each gene. Figure 6.9 shows one possible initial population of size $n = 4$.

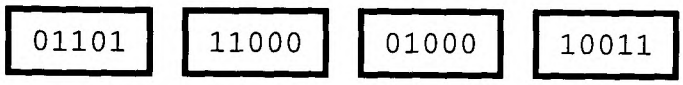


Figure 6.9 An initial population of four chromosomes.

Starting with this initial population, the mating pool for the next generation is selected using a biased roulette wheel. Table 6.2 shows each chromosome, its associated fitness value and the relative size of its slot on a weighted roulette wheel. The Table also indicates the actual count of occurrences for each chromosome resulting from 4 spins of the wheel. Figure 6.10 shows the weighted roulette wheel and the relative probabilities of each chromosome being selected during each spin.

Index	*Chromosome*	*Value (x)*	*Fitness (x^2)*	*% of total*	*Count*
A_1	01101	13	169	14.4	1
A_2	11000	24	576	49.2	2
A_3	01000	8	64	5.5	0
A_4	10011	19	361	30.9	1
Sum			1170	100	4
Average			293	25	1
Max			576	49.2	2

Table 6.2 Example of a simple GA: initial population of chromosomes and their fitness values.

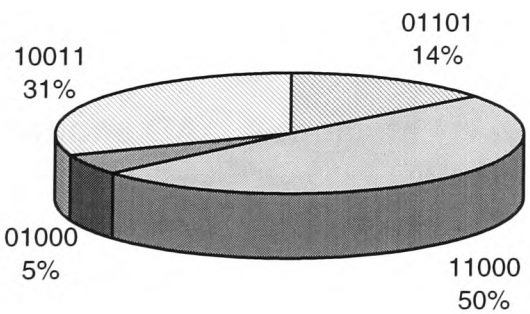


Figure 6.10 Size of slots of weighted roulette wheel for the initial population (rounded upwards to the nearest integer).

Table 6.2 indicates that the mating pool for the initial population will contain one copy each of chromosomes A_1 and A_4, two copies of chromosome A_2 and no copies of chromosome A_3. Chromosomes are randomly paired for mating and crossover sites are randomly selected for each pair, as shown in Table 6.3. In this example, the crossover probability is 1.0 meaning that none of the chromosomes from the previous generation survives unchanged into the next.

Index	*Mating pool and crossover site*	*Index number of mate*	*Crossover locus*	*New population after reproduction*	*Value (x)*	*Fitness (x^2)*
A_1	0110 | 1	2	4	01100	12	144
A_2	1100 | 0	1	4	11001	25	625
A_3	11 | 000	4	2	11011	27	729
A_4	10 | 011	2	2	10000	156	256
Sum						1754
Averag						439
Max						729

Table 6.3 Example GA: second generation of chromosomes.

The last operation, mutation, is usually performed on a bit-by-bit basis. The mutation probability in this example was set to 0.001. Therefore, with 20 transferred bit position at each generation, one would expect $20 \times 0.001 = 0.02$ bits to undergo mutation in each generation. Consequently, for our single generation example, no bits have undergone mutation. Whilst drawing conclusions from a single iteration of a stochastic process is risky, Table 6.3 indicates that in our example both the maximal and average fitness have improved in the second generation of the population.

6.5.2. Foundations GAs: Schemata

In the simple genetic algorithm described above the only information extracted from the chromosomes were their fitness values. However, by examining the alleles of each chromosome certain similarities can be found: certain bit patterns seem to be associated with high fitness values. For example, chromosomes starting with a 1 seem to be amongst the best! Searching for similarities between chromosomes of high fitness introduces a wealth of new information to assist in the optimization (Goldberg and Deb 1991).

A schema is a similarity template describing a subset of chromosomes with similarities amongst certain of their alleles. In the example above alleles were extracted from the binary alphabet, $V = \{0,$

1}. Schemata operate on an extended alphabet by introducing a wild card allele *, which matches any allele from the alphabet. In our above example schemata would operate on the extended alphabet $V+$ = {0, 1, *}. The schema H = [*111*] matches four chromosomes, [01110], [01111], [11110] and [11111]. Consider a set of chromosomes of length l. For an alphabet of cardinality k, the use of schemata increases the number of possible arrangements of alleles (and therefore the size of the search space) from k^l to $(k+1)^l$, however by considering similarities new information is introduced to assist the search.

A single chromosome of length l is a member of 2^l schemata, because each gene may take on its actual value or the wild card character. A population of size n_p will therefore, contain between 2^l and $n_p \times 2^l$ schemata depending upon the population diversity.

Some schemata are more specific than others, i.e. they contain fewer wild card characters: the schema [011*1**] is more specific than [0******]. Moreover, some schemata span more of the length of the chromosome than others, e.g. [1****1*] spans a longer portion of the chromosome than [1*1****]. The *order* of a schema H is the number of non-wild card alleles in the schema and is denoted by $o(H)$. The *defining length* of a schema is defined as the distance between the first and last fixed locus of the schema, and is denoted by $\delta(H)$:

H = [011*1**] $\Rightarrow$ $o(H) = 4$ $\Rightarrow$ $\delta(H) = 4$

H = [0******] $\Rightarrow$ $o(H) = 1$ $\Rightarrow$ $\delta(H) = 0$

Figure 6.11 The order and defining length of two schemata.

These properties of schemata have a great significance during reproduction. The crossover operator acts on a schema, the schema will not be disrupted if the crossover locus does not dissect the defining length of the schema. Consider the following chromosome A, undergoing reproduction, and two schemata, H_1 and H_2:

A = [011 | 1000]

H_1 = [*1* | ***0] $\Rightarrow$ $o(H) = 2$ $\Rightarrow$ $\delta(H) = 5$

H_2 = [*** | 10**] $\Rightarrow$ $o(H) = 2$ $\Rightarrow$ $\delta(H) = 1$

Figure 6.12 A chromosome, A and two schemata, H_1 and H_2 prior to undergoing reproduction with a crossover locus of 3, indicated by the dotted line.

Suppose that when A is mated to another chromosome (B), the crossover locus is 3, as indicated by the dotted line in Figure 6.12. Unless B is identical to A at the fixed positions of the schema, H_1 will be destroyed during reproduction. Genes on either side of the crossover locus are placed in different offspring. Therefore, in Figure 6.12, the `[1]` at position 2, and the `[0]` at position 7 of H_1 will be placed in different offspring, destroying the schema unless B is identical to A at these loci. Similarly, $H2$ will survive crossover because the fixed positions of the schema (positions 4 and 5) both lie on one side of the crossover locus and therefore will both be placed, intact, into the same offspring. Although the crossover locus may be at any location along the chromosome, it can be seen that H_1 is less likely to survive crossover intact than H_2 because, on average, the cross-over locus is more likely to fall between the fixed gene positions. To quantify this observation we note that if the defining length of a schema is $\delta(H)$, and that if the crossover locus if selected randomly from the $(l - 1)$ possible sites, then the probability p_d, of the schema H being destroyed during crossover is (Forest and Mitchell 1995):

$$p_d = \frac{\delta(H)}{(l-1)} \tag{6.36}$$

From Equation 6.36, the probabilities of schemata H_1 and H_2 being destroyed are 5/6 and 1/6 respectively. Mutation is normally set to quite low rates of probability and consequently does alter a given schema very frequently (Vose and Liepins 1991). Therefore, highly fit, short length schemata have a high probability of propagating into the next generation.

It is useful to make mathematically explicit the expected number of schemata in the population at each generation. Suppose that at a given iteration, t of the GA, there are $m = \mathrm{m}(H, t)$ examples of schema H, contained within a population $\mathbf{A}(t)$ of size n. During reproduction a chromosome A_i is selected with a probability given by:

$$p_i = \frac{f_i}{\sum_{j=1}^{n} f_j} \tag{6.37a}$$

Where: f_i is the fitness of the i^{th} chromosome.

Alternatively, the lower bound on the crossover survival probability of a schema, p_s, may be calculated. A schema survives when the crossover locus falls outside the defining length, so for a crossover probability of p_c, the survival probability p_s is bound by:

$$p_s \geq 1 - p_c \times \frac{\delta(H)}{l-1} \tag{6.37b}$$

At the next iteration, the number of examples of H will be $m(H, t+1)$, given by:

$$\mathrm{m}(H,t+1) = \mathrm{m}(H,t) \times n \times \frac{\bar{f}(H)}{\sum_{j=1}^{n} f_j} \tag{6.38}$$

Where: $\bar{f}(H)$ is the average fitness of the chromosomes representing the schema H.

The average fitness of the entire population can be written:

$$\bar{f} = \frac{\sum_{j=1}^{n} f_j}{n} \tag{6.39}$$

Combining Equations 6.38 and 6.39:

$$\mathrm{m}(H,t+1) = \mathrm{m}(H,t) \times \frac{\bar{f}(H)}{\bar{f}} \tag{6.40}$$

Equation 6.40 states that the growth rate of a schema under the operation of reproduction is given by the ratio of the fitness of the schema to the fitness of the population. Therefore, schemata with above average fitness will have an increasing number of examples in the next generation, whereas schemata with fitness values below the population average will receive a diminishing number of examples. This behaviour will occur for all schemata H in the population in parallel.

Combining reproduction and crossover using Equations 6.37 and 6.40, we obtain:

$$\mathrm{m}(H,t+1) \geq \mathrm{m}(H,t) \times \frac{\bar{f}(H)}{\bar{f}} \left[1 - p_c \times \frac{\delta(H)}{l-1} \right] \tag{6.41}$$

Equation 6.41 defines the minimum number of a given schemata surviving into the next generation, under the operations of both reproduction and crossover. The difference between Equation 6.41 and 6.40, the equation for reproduction alone, is the presence of the crossover survival term on the right hand side. Under both reproduction and crossover, survival depends upon whether the schema has above or below average fitness and whether it has a relatively short or relatively long defining length.

Those schemata with both above average fitness and short defining lengths will be sampled at exponentially increasing rates in subsequent iterations.

Mutation has little impact on the situation so far described. If the probability of the random alteration of a single gene is p_m, then the probability of a single allele surviving mutation is $(1 - p_m)$ and a given schema survives if each of its o(H) fixed positions survives. Therefore the survival probability of the schema is:

$$(1-p_m)^{o(H)} \text{ and for } p_m << 1 \quad \Rightarrow \quad 1-o(H)\times p_m \tag{6.42}$$

Finally combining Equations 6.41 and 6.42, and simplifying, the number of examples of H surviving into the next generation is :

$$m(H,t+1) \geq m(H,t)\times\frac{\bar{f}(H)}{\bar{f}}\left[1-p_c\times\frac{\delta(H)}{l-1}-o(H)\times p_m\right] \tag{6.43}$$

Equation 6.43 is the foundation of genetic algorithms and is known as the *Schema Theorem.*

By extending the simple GA example above (Figure 6.8), the effect of reproduction on three schemata can be illustrated. Consider three schemata: H_1 = [1****], H_2 = [*10**] *and* H_3 = [1***0].

Before reproduction				*After reproduction*	*After all operations*
Index	Schema	Chromosome representation	Schema Average fitness f(H)	Chromosome representation	Chromosome representation
H_1	1****	A_2, A_4	469	A_2, A_3, A_4	A_2, A_3, A_4
H_2	*10**	A_2, A_3	320	A_2, A_3	A_2, A_3
H_3	1***0	A_2	576	A_2, A_3	A_4
A_1	01101			01101	01100
A_2	11000			11000	11001
A_3	01000			11000	11011
A_4	10011			10011	10000

Table 6.4 An example of schemata processing.

For schema H_1 two chromosomes are initially represented by this schema, A_2 and A_4, and after reproduction, the number of represented chromosomes increases to three, as predicted by the schema theorem:

$$m(H,t)\times\frac{\bar{f}(H)}{\bar{f}} \Rightarrow 2\times\frac{\left[\frac{576+361}{2}\right]}{293} \Rightarrow 3.2$$

Because the defining length of the schema is 0, crossover can have no effect on it (as seen in Table 6.4), and if the mutation probability is set to a typically low value of 0.001, the probability of the single fixed position of the schema changing in any of the three chromosome representations is only 0.003. As a result the number of representations of H_1 after all operations remains 3. Schema H_2 has two representatives both before and after reproduction, once again agreeing with the schema theorem. The short defining length of H_2 means that there is a relatively high probability of the schema surviving crossover (from Equation 6.36, $p_d = ¼$) and as indicated in Table 6.4, this is the case in this example. Schema H_3 starts with a single representative and has two following reproduction, in accordance with the schema theorem. However, the long defining length of this schema means that crossover usually destroys it. This is illustrated in the example by only one representative remaining after all operations.

Summary

Genetic algorithms are a very powerful optimization and search tool. The three basic GA operators are very simple. involving nothing more than random number generation, string copying and partial string exchanging. The searching strategy of GAs is completely blind: searching is stochastic process that does not use deterministic rules but is performed by sampling alone (Turner 1987). Unlike traditional optimization heuristics, GAs generally consider only the value of the merit function, and other ancillary information is ignored.

6.6. COMPARISON OF OPTIMIZATION TECHNIQUES

The performance of each of the 4 optimization techniques, downhill simplex, direction set, simulated annealing and genetic algorithms, was compared by attempting to maximize[5] a particularly nasty n-dimensional function given by:

$$f(x_i)=\frac{\text{abs}\left[\sum_{i=1}^{n}\cos^4(x_i)-2\prod_{i=1}^{n}\cos^2(x_i)\right]}{\sqrt{\sum_{i=1}^{n}ix_i^2}} \tag{6.44}$$

[5] Note that the processes of maximization and minimization are trivially related.

$$\text{with: } 0 < x_i < 10; \quad i = 1,\ldots,n; \quad \prod_{i=1}^{n} x_i > 1.8; \quad \sum_{i=1}^{n} x_i < 30\frac{n}{2}$$

This function, introduced by Keane (1995), is widely used for testing optimization algorithms because it produces a very bumpy surface, as shown in Figure 6.13. The series of peaks gets smaller with distance from the origin, and is nearly, but not quite, symmetrical about $x_1 = x_j$ so that peaks always occur in pairs, with one slightly bigger that its twin. The global maximum of the function for n = 2, is 0.26 at (3.087, 1.517). A function with so many separate different peak of similar, but increasing height is extremely difficult to maximize because so many local maxima exist.

Computer programs were written in C++ to implement the four optimizations algorithms. The software additionally was designed to carry out integrity checking and timing of the individual optimization runs. The code was highly optimized and compiled using SUNC++ 4.0.1 compiler. It was run on a variety of Sun Microsystems hardware under Solaris 2.4 and 2.5 series operating systems. Timed runs were executed on a four-processor Ultra-SPARC server[6].

The centre of the feature space, with $x_1 = 5$, was chosen as the starting point for the optimization runs. At this point, the value of the function is always zero. An initial set of opimization runs was performed for $n = 2$. The constraints were applied by setting the function value to an extremely small number if the constraints were violated. For both the downhill simplex and direction set routines, the maximum number of iterations were set to 1000, with a maximum of 10 restarts.

Parameter	*Value*
K	1000
P	0.3
W	5.0
C	2.0

Table 6.5 Annealing schedule parameters

[6] Object code was recomplied to run "native" on Ultra-SPARC hardware.

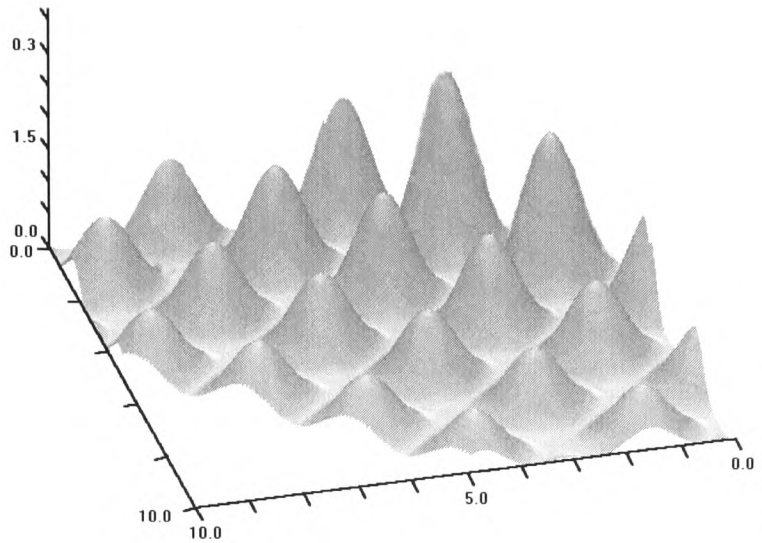


Figure 6.13 The "bump" function for $n = 2$.

The simulated annealing algorithm used was SA with downhill simplex search, and the logarithmic annealing schedule, with the schedule parameters given in Table 6.5. The parameters for the GA are given in Table 6.6.

Parameter	*Value*
Number of generations (n_p)	10
Population size	50
Survival probability (p_s)	0.8
Crossover probability (p_c)	0.8
Mutation probability (p_m)	0.0004
Elitism	TRUE

Table 6.6 Genetic algorithm parameters for the bump function with $n = 2$.

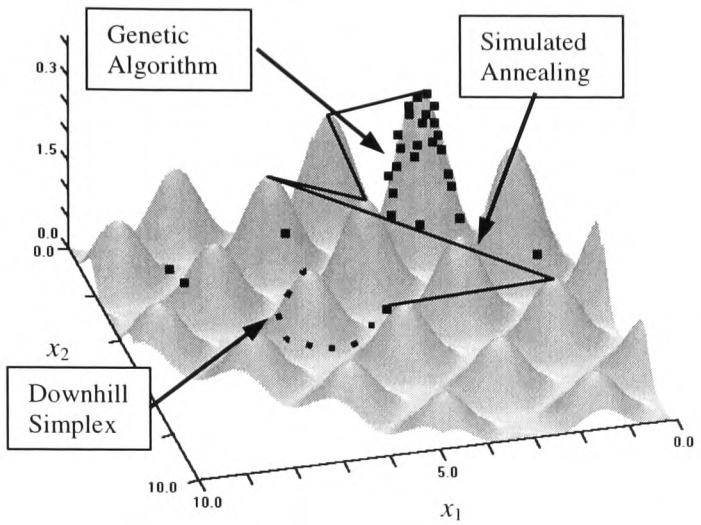


Figure 6.14 Results of optimization of the "bump" function for $n = 2$, dotted line – downhill simplex, soilid line – simulated annealing, black squares genetic algorithm. (Only points from the final generation and annealing temperature are shown.)

Figure 6.14 shows the results of the optimization runs. The direction set method continually became trapped along the constraints of the problem and failed to return satisfactory results in any run. This suggests either that the problem is too tough for this technique to handle, or that a more sophisticated constraint application procedure is required in order to use the direction set algorithm. The downhill simplex algorithm quickly converged to a nearby peak, thus finding a local, not global maximum. In contrast, both the SA and GA techniques show a good performance after 1000 iterations, both finding very close approximations to the global maximum. SA and GA are stochastic algorithms and it is important to know the consistency of performance of these techniques. Consequently, the runs for SA and GA were repeated using different random number sequences; the results are shown in Figure 6.15. As can be seen, variations in the result of the optimization can occur when different random number sequences are used. Therefore, with stochastic optimization methods such as SA and GA it is advisable to take the average of a number of shorter runs rather than the result of one long one.

The optimizations runs were repeated for the "bump" function with $n = 50$. This is an extremely severe test, and consequently the maximum number of iterations for each of the four techniques was set to 150,000. The GA population size was increased to 250. To overcome the problems encountered with the direction set algorithm described above, the constraint application method was altered. Constraints were applied setting the function values to a low value proportional to the distance moved beyond the constraint boundary, with an additional incremental penalty added for repeated excursions beyond the constraining limits.

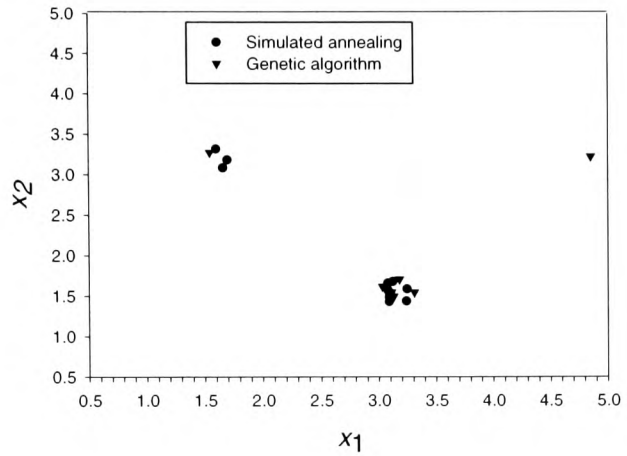


Figure 6.15 Results of ten SA and GA optimizations runs of the "bump" function, for $n = 2$ with different random number sequences.

Table 6.7 compares the results obtained using the four techniques. As this table illustrates, SA and GA significantly outperform the non-stochastic techniques. The maxima for downhill simplex and direction set where reached when the algorithms terminated after approximately 3000 iterations. After the same number of iterations, the stochastic techniques had achieved similar values for the maximum, but were able to continue searching for another 147,000 iterations. The computing time required for 150,000 iterations of the stochastic techniques is considerable, taking many hours even on high performance computers. This issue will be addressed further in the next chapter.

Method	*Maximum*
Genetic algorithm	0.767
Simulated annealing	0.523
Direction set	0.286
Downhill simplex	0.291

Table 6.7 The results for optimization of the "bump" function for $n = 50$. Averages taken over 5 runs.

Simulated annealing parameter tuning

The results also show that the genetic algorithm runs significantly out performed simulated annealing, with the default controlling parameters. Tuning the controlling parameters can significantly improve

the performance of the GA or SA optimization. Unfortunately, parameter tuning is an optimization problem it its own right.

The performance of SA can be strongly dependent upon the choice of annealing schedule parameters, therefore a poor choice of annealing parameters may result in only a sub-set of the feature space being searched.

Unless unlimited computing resources are available, it is normal to fix the maximum number of iterations (K) allowed for the procedure (Equations 6.35a,b,c). The choice of annealing schedule parameters will depend upon the nature of the optimization problem. For example, if the feature space is characterized by deep, well separated, local minima, then a high initial temperature is required to enable the algorithm to jump out of these local minima into new regions. The more difficult the optimization problem the greater the number of iterations that will be required. There are three factors that control the effectiveness of the annealing schedule:

- The initial temperature T_0
  If T_0 is too hot, the algorithm becomes computationally inefficient, spending too much time making "uphill jumps".
  If T_0 is too cold, the system is unable to jump out of local minima and only a sub-section of the feature space is searched.

- The rate of cooling - the number of function evaluations at each temperature
  If cooling is too fast and the system is quenched into an amorphous state.
  If cooling is too slow, the algorithm is inefficient.
  This is related to the number of function evaluations allowed as each temperature.
  If too many evaluations are allowed at a given temperature then the number of temperature steps must be reduced resulting in quenching.
  If too few evaluations are allowed then the search become inefficient.

- Bias of iterations with temperature.
  If too many iterations are used at high temperatures then final cooling may be too rapid to find the global minimum.
  If too few iterations are used at high temperatures then only a sub-section of the feature space may be searched.

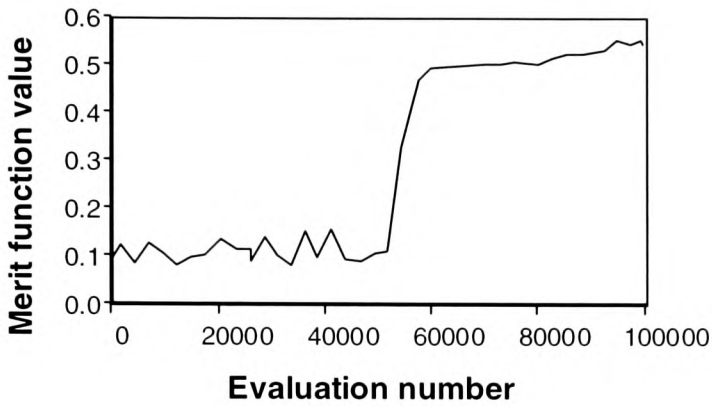


Figure 6.16 Merit function trace for simulated annealing (SA2) optimization of the "bump" function, with the logarithmic schedule (K = 100,000, P = 0.3, W = 5 and C = 2)

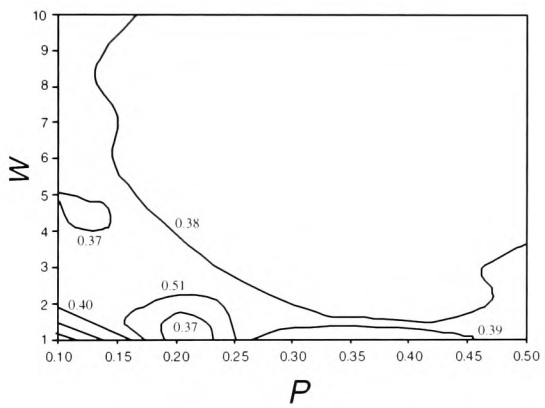


Figure 6.17 The effect of varying two of the annealing schedule parameters on SA optimization of the "bump" function. (C = 2)

For the logarithmic annealing schedule the initial temperature is determined implicitly by the three parameters P, W and C, and by the maximum number of iterations K. For the bump problem, K was set to 100,000. Setting the other parameters to their default values (P = 0.3, W = 5 and C = 2), results in an initial temperature in excess of 10^{16}, and results in more than 2100 evaluations at each temperature, with the temperature being reduced in 46 steps. The optimization trace for the bump function using these parameters is shown in Figure 6.16. In this figure, the liquid freezing and frozen stages are clearly visible, indicating that the initial temperature was hot enough, and that the combinations of, the number of iterations per temperature step, and the number of temperature steps, was sufficient to achieve annealing. However, were the parameters chosen the optimum ones? In Figure 6.16 is can be seen that the greatest improvements in the merit function occur at iterations in

the middle of the range. This, combined with the long initial period with little improvement in the merit function, suggests that the initial temperature was too high. To investigate whether the optimization performance could be improved by modifying the schedule, optimization runs were performed whilst the parameters W and P were varied, and C was kept fixed. Figure 6.17 shows a contour map of the variation in the value of the bump function with the two annealing schedule parameters, P controlling the number of trails at each temperature, and W controlling the range of temperatures. The figure shows that the performance of SA1 does not improved markedly, even for quite broad changes in the schedule parameters, and does not approach the excellent results produced by GA. However, the schedule does give consistent results over a wide variation of parameter values. The performance with "bump" seems best for values of W around 2.0, and a low value of P around 0.2, i.e. a few distinct temperatures, giving a very much cooler initial temperature of 32 compared to the default.

For the alpha schedule, the value of T_0 is critical; if it is set too low then the algorithm may never escape from the first local optimum it encounters. This is illustrated in Figure 6.18 which shows SA2 optimization of the bump function with $T_0 = 100$ and $K = 100{,}000$. One way to avoid this is to set T_0 using the specific heat. With this method, T_0 was typically found to take values of around 5000 which is in line with the results of Figure 6.17. The number of iterations per temperature sets f_{lev} is a large fraction of the total number of steps K, then the resulting schedule performs a relatively large number of iterations a small number of different temperature levels. This seems to work well for the bump functions, but for other problems works less well (Chapter 7).

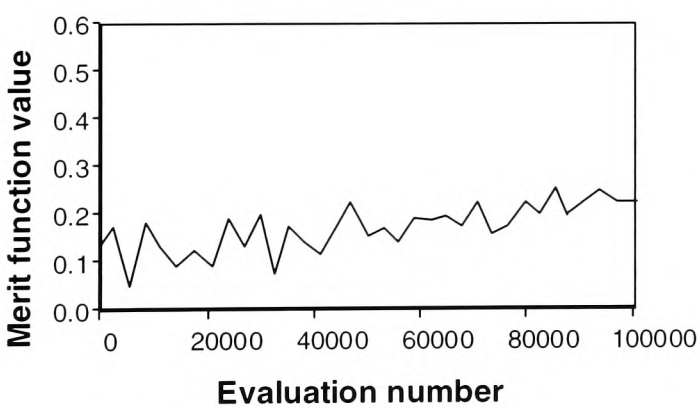


Figure 6.18 Merit function trace for simulated annealing (SA2) optimization of the "bump" function, illustrating a low initial temperature. ($K = 100{,}000$, $T_0 = 100$, $f_{lev} = 1000$, $\alpha = 1$)

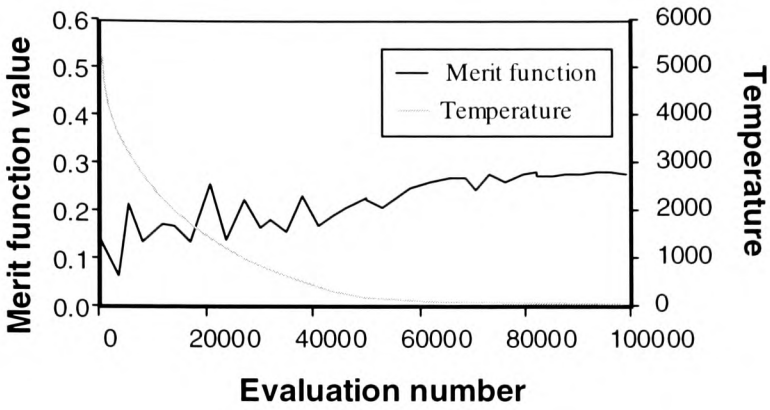


Figure 6.19 Merit function trace for simulated annealing (SA2) optimization of the "bump" function, illustrating the effect of quenching. (K = 100,000, T_0 = 5100, f_{lev} = 100, α = 4).

The temperature must be reduced sufficiently slowly for annealing to take place. If the reduction in temperature is too rapid, quenching results and the systems tends to become frozen into a local optimum. This is illustrated in Figure 6.19, where a value for α of 4, causes a rapid early cooling. Attempts by researchers to find a problem-independent method for selection of the optimum combination of annealing parameters has not been found. Each problem has its own characteristics that favour one combination or another, thus trial an error remains the best tool available for this purpose. Table 6.8 shows the "optimum" set of schedule parameters found by trail an error for the bump problem using SA2. The values of the merit function at 7.32 is still not as good as the results from GA, but is nevertheless a significant improvement of the default schedule for SA1.

Parameter	*Value*
T_0	5000
K	120000
f_{lev}	100
α	1.0
Merit function	0.712

Table 6.8 "Optimum" set of annealing schedule parameters, and the corresponding merit function value for the bump problem using SA2.

6.6.1. Conclusion

Stochastic optimization techniques such as genetic algorithms and simulated annealing offer significant advantages over traditional search techniques such as downhill simplex and direction set for difficult functions (e.g. the bump function) where many local minima or maxima exist. To get the best out of the simulated annealing technique, the annealing schedule and control parameters need to be hand coded for each problem. Improvements can also be made to the performance of GA by tuning

parameters, or modifying the basic algorithm or encoding, however, these improvements are small and GA is significantly less sensitive to its control parameters than SA. These results are in agreement with much of the recent work in this field, e.g. Fogel (1993). Both GA and SA exhibit some sensitivity to random number sequences. Therefore, for maximum confidence in the results from these methods, it is necessary to take averages over a number of runs. The main limitation on the use of stochastic optimization techniques is the large computational burden involved, which necessitates the employment of high performance computers.

Chapter 7

Application of Optimization Techniques

In Chapter 6 four optimization techniques were described and the performance of each technique for optimizing a difficult but abstract function compared. In this chapter, the four techniques are applied to the optimization of SVAT model components. As outlined in Chapter 5, there are two main ways in which remotely sensed data can be integrated into SVAT models. Firstly, by inverting models that predict remote sensing signals (e.g. BRDF models) to determine vegetation parameters, and to use these to parameterise the SVAT model. Secondly, by optimizing components of the SVAT model itself to constrain its behaviour within limits imposed by remote sensing measurements. This Chapter is concerned with the former and Chapter 8 with the latter.

The potential for inversion of two canopy reflectance models was investigated. The two models used in this study, SAIL and Li-Strahler, differ markedly in their approach and intended application (see Chapter 3). The SAIL model takes the turbid medium approach (Chapter 3) and assumes a horizontally homogeneous canopy. It therefore best represents the BRDF of crop canopies such as wheat. The Li-Strahler model adopts a geometrical-optical approach (Chapter 3) and is specifically designed for modelling the reflectance of forest canopies.

7.1. GENERATION OF RANDOM NUMBERS

Stochastic optimization techniques such as genetic algorithms or simulated annealing make extensive use of random numbers in their algorithms. As shown in Chapter 6, the choice of random number sequence may affect the result of an optimization using stochastic techniques. Clearly the method of generating random numbers is important. Computers and computer programs are, by necessity deterministic (or at least the programmers hope they are), it is not necessarily straightforward, therefore, to produce a non –deterministic series of numbers from a deterministic program. Nevertheless, random number sequences are so important in many areas of science that a considerable body of work on random number algorithms exists.

Applications may differ in their required random number distribution. The most commonly required is one where numbers are generated from within a specific range of values (e.g. 0 to 1) where every number is just as likely to occur as every other number. Such numbers are known as *uniform deviates.*

In contrast, Gaussian deviates are drawn from a normal distribution with a specified mean and standard deviation. Computer algorithms usually derive such more complex distributions from a set of uniform deviates and so the reliable generation of uniform deviates forms the basis of a random number generating program.

Random number generators are built into many compiler/operating systems, e.g. the function `rand()`, in Sun C under Solaris. This function returns an integer in the range 0 to a system specific maximum number, RAND_MAX. The function is initialized with a *seed*, and will then produce a series of pseudo-random numbers from a very long sequence. The same initializing value for *seed* will always produce the same sequence of numbers. The commonly used algorithm for generating random numbers is:

$$I_{j+1} = \{aI_j + c\} \bmod \{m\} \tag{7.1}$$

where: I_j is the jth random number in the sequence

a is a positive integer, known as the multiplier

c is a positive integer, known as the increment

m is known as the modulus and is given by m = (RAND_MAX – 1)

Linear congruential generators such as that given by Equation 7.1 produce random number sequences that are quite acceptable for many simple problems and because they involve few operations, they are very fast. However, they may be flawed for use in demanding applications such as stochastic optimization. Obviously a, c and m must be chosen carefully, so that every number between 0 and the maximum, m will appear at some point. The choice of the initial seed, I_0, merely determines the starting point of the series along the sequence. However, the function `rand()` returns an integer, which is only held as a two-byte value on many machines. Moreover, the value of RAND_MAX on many systems is not very large – the ANSI standard suggests m = 32767 which does not give a very large sequence of random numbers before the series repeats. The fact that linear congruential generators produce series of numbers that are not free from sequential correlation on successive calls is a major disadvantage for their use in stochastic optimization. This is because when the generator is used to create co-ordinates in n-dimensional space, the points will tend to lie on (n - 1) dimensional planes rather than occupying the whole of the space; there will be at most $m^{1/k}$ such planes, (Park and Miller, 1988). However, Park and Miller (1988) show that a simplified version of Equation 7.1 with c = 0, can perform as well as many more complex random number algorithms for specific values of a, c and m, shown in Table 7.3

Parameter	*Value*
a	$7^5 = 16807$

m	$2^{31} - 1 = 2147483627$
c	0

Table 7.1 Parameters for Equation 7.1, according to Park and Miller (1988)

Unfortunately, direct implementation of Equation 7.1 with the parameters of Table 7.1 is not possible in a language such as C because $a \times (m - 1)$ exceeds the maximum value for a 32 bit integer. Knuth (1981) describes a simple algorithm to avoid this problem.

Let: $q = \text{integer part}\{m/a\}$, $r = \{m\} \bmod \{a\}$ so that $m = aq + ri$

For $r < q$ and $0 < I < (m - 1)$ it can be shown that

$$0 < a[\{z\} \bmod \{q\}] < (m - 1)$$

$$0 < r(z/q) < (m - 1)$$

and

$$a[\{z\} \bmod \{q\}] = \begin{cases} a[\{z\} \bmod \{q\}] - r(z/q) & \geq 0 \\ a[\{z\} \bmod \{q\}] - r(z/q) + m & \text{otherwise} \end{cases} \qquad (7.2)$$

Using $m = 2^{31}$, gives $q = 127773$ and $r = 2836$, using the parameters of Table 7.1. Equations 7.1 and 7.2 from the *Minimal Standard* random number generator (Park and Miller, 1989). Although the minimal standard is sufficient for many applications, it has some disadvantages. Zero must never be used as the initial seed as it perpetuates itself: conversely, zero never occurs for a non-zero initial seed. This problem can be avoided by some skilful programming, but other more subtle problems require refinement of the basic algorithm. L'Ecuyler (1989) illustrates that there are low-order serial correlations present in the minimal standard generator that result in it failing the χ^2 test[7] in certain circumstances[8], and a further problem is that very small random numbers generated by the algorithm tend to be followed by below average values. Knuth (1981), proposed a modification to the minimal standard to remove low-order serial correlations by shuffling the order that deviates occur in the sequence, as shown in Figure 7.1.

[7] The χ^2 test is : $\chi^2 = \sum_i \frac{(R_i - S_i)^2}{R_i + S_i}$ where R_i and S_i are the number of events in bin i of first and second set respectively.

[8] Successive random numbers (R_i, S_i), binned into a 2-dimensional plane, fail the χ^2 test, for $N \approx 10^7$.

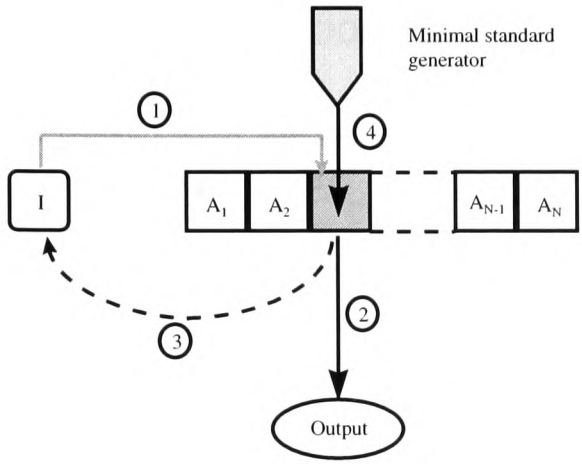


Figure 7.1 Shuffle improvement of the minimal standard random number generator.
(1) The random number, I, is used to indicate an index position in the shuffle table A, of size N, which has been pre-loaded with random numbers from the minimal standard generator.
(2) The random number A_I, becomes the next output number in the sequence.
(3) A_I also becomes the next value for the index I.
(4) The minimal standard generator replaces A_I with a new random number.

The shuffled minimal standard generator passes all statistical tests until the number of random numbers returned approaches the order of the period m. For very long sequences of random numbers, L'Ecuyer (1989) proposed a simple way of combining two (or more) different sequences, with different periods to obtain a new sequence whose period is at least the multiple of the lengths of the two individual sequences. An obvious implementation might be to add the values of the two sequences modulo the modulus of either of the them, say m. This, at first sight, seems a trivial task, however, intermediate values exceeding the maximum values for an integer must be avoided. L'Ecuyer's method subtracts the values of the two sequences and if this produces a negative number adds $(m - 1)$. This clever trick always produces numbers in the range $0 \ldots m - 1$. The parameters used by L'Ecuyer (1989) given in Table 7.2 produce a sequence with a period of larger than 10^{18}, which is easily sufficient for even the most intensive application. Therefore this technique was implemented and used throughout this study.

Parameter	*Value*	*Parameter*	*Value*
a_1	40014	a_2	40692
m_1	2147483563	m_2	2147483399
q_1	53668	q_2	52774
r_1	12211	r_2	3791

Table 7.2 Parameters used by L'Ecuyer (1989) in his combined sequence generator.

7.2. INVERSION OF THE SAIL MODEL USING DIRECTIONAL REFLECTANCE DATA

The SAIL model (Verhoef, 1984) is a widely used BRDF model because of its simplicity and good performance for homogeneous canopies (Goel, 1987). In order to study the inversion of BRDF models, numerous accurate reflectance measurements at a wide range of solar and view zenith and azimuth angles must be combined with near-simultaneous detailed measurements of canopy and soil physical and optical properties.

Detailed BRDF datasets for homogeneous canopies such as wheat are not widely available. In fact, the Kimes dataset (Kimes *et al.* 1985) represents, to the author's knowledge, the only BRDF dataset for wheat, and datasets for other crops are equally uncommon. In recent years research has concentrated on the measurement and modelling of forest BRDF. Accurate measurement and modelling of the worlds forests are vitally important for understanding the likely impacts of climate change on vegetation and the possible feedback effects which may occur between vegetation and the climate. The complex three-dimensional structure of forest canopies makes accurate modelling of their BRDF difficult. Unfortunately, complex BRDF models capable of representing the forest BRDF accurately tend to have large numbers of parameters which preclude their inversion (Nilson *et al.* 1995).

Unfortunately, the Kimes dataset is not suitable for rigorous testing because it lacks canopy physical data. For the two types of wheat crop considered, hard wheat and irrigated wheat (Kimes *et al.* 1985), measurements of LAI and LAD were not made: I therefore derived LAI using equations from Kimes (1974) assuming the leaf angle distribution was erectophile. Measurements of leaf optical properties and soil reflectance were not reported, making this dataset useless for inversion studies.

Major *et al.* (1992) inverted the SAIL model using maize reflectance data, to obtain estimates of LAI. Nadir reflectance of six maize genotypes was measured for different solar zenith angles in red and near-infrared wavebands. Measurements of LAI and soil reflectance were made, but leaf optical properties and LAD were obtained from the best fit between SAIL modelled reflectance and measured data. Two variants of the SAIL model were inverted: (i) the basic SAIL model (Verhoef, 1984), and (ii) TRIM (Goel and Grier, 1988) – a modified version of SAIL to include row effects - see Chapter 3. The models were inverted by using them to calculate the Transformed Soil-Adjusted Vegetation Index (TSAVI), V_{TSAVI}, (Baret *et al.* 1989) for a number of different LAI values. TSAVI is related to LAI, L, and the extinction coefficient, k, by the equation:

$$L = \frac{-1}{k \ln(1 - V_{TSAVI})} \tag{7.1}$$

Equation 7.1 was fitted to the SAIL generated TSAVI data and the derived value of k, to determine the LAI. Results for the basic SAIL model showed that modelled reflectance reached saturation before

measured data in both red and NIR wavebands indicating the influence of bare soil reflectance in the measured signal. This resulted in an underestimation of LAI (averaged over all measurements) of 37.6%. The TRIM variation of the SAIL model performed much better, overestimating LAI by 8.3%, which is of the same order as the typical error of field measurement (see Chapter 2). The SAIL model assumes a horizontally homogeneous canopy and it is therefore unsurprising that it does not perform well for crops that exhibit a distinct row structure because of typical sowing practices. However, the relatively good LAI results obtained by inverting the row-modified SAIL model by Major *et al.* (1992) is encouraging.

Ranson *et al.* (1985) made extensive canopy biophysical and optical measurements in addition to BRDF measurements for a soybean canopy in Indiana, USA. These included LAI, LAD, canopy height, leaf and soil optical properties in four wavebands, together with limited incident radiation data, for three dates (July 17, July 24 and August 27) during the growing season in 1980. The three dates chosen also offered a range of different canopy coverage, with complete canopy closure occurring for the latter date, August 27. This dataset is one of the few available for "homogeneous" canopies and consequently has been widely used in inversion studies, notably in a series of papers by Goel *et al.* (Goel and Grier 1986a,b, Goel, Strebel and Thompson 1985, Goel and Thompson 1984b,c), the main drawback with the dataset being the limited number of wavebands used. Goel and Thompson (1984c) inverted the SAIL model on the Ranson dataset (NIR only) using a least squares fitting approach with a number of modifications including the use of specific angle transforms (Goel, Strebel and Thompson 1984, Goel and Thompson 1984a,b). They were unable to achieve acceptable results when all model parameters were free during inversion but obtained good LAI estimations when all but the LAI and LAD parameters were fixed. In their study, Goel and Thompson, (1984c) found that there were significant systematic differences between measured canopy reflectances and those modelled by SAIL when parameterised with the measured parameter set, but the fit between measured reflectances and modelled values when SAIL was parameterised with the parameters obtained from the inversion was significantly better. The probable reason for this is a combination of inadequacies in the ability of SAIL to model the BRDF of soybeans, errors in measurements of both canopy parameters and reflectances, the presence of canopy components other than leaves, e.g. pods and stems, the lack of a hot-spot representation in SAIL, and underlying inhomogeneities in the soybean canopy introducing "row effects". The HOTSAIL and TRIM (Goel and Grier 1988) models address the problems of the hot-spot and row effects respectively.

7.2.1. Simulated Data

Because of the scarcity of suitable datasets for testing the optimization procedures, optimizations were initially carried out using *simulated* data. Firstly SAIL was run with a given set of input parameters to calculate canopy directional reflectances, then these reflectances were used as the canopy reflectance

measurements. The input parameters were "forgotten", initial guesses were made for the input parameters and the optimization procedure started.

The inversion of the SAIL model with no *a priori* knowledge requires the determination of seven parameters (LAI, two parameters to express the leaf angle distribution, leaf reflectance and transmittance, soil reflectance, and the fraction of diffuse to total incident radiation), as shown in Table 7.4, and hence require a seven dimensional minimization. The other four parameters, the solar zenith angle, the viewing zenith and relative azimuth angles, and the wavelength of the observation are determined by the sun-observer geometry, and hence are fixed for a given reflectance measurement.

Although several researchers have reported successful inversions of the SAIL model, this has been with a large number of measured directional data (Goel and Thompson, 1984b) and only a few parameters free in the inversion process. The maximum number of directional measurements that can be made by a satellite sensor is restricted by the physics of satellite orbits and the engineering skills required to build the instrument itself. One of the most advanced of the current high resolution sensors, ASAS, carries an instrument with seven looks (Table 7.3).

Number of spectral bands	30
Spectral sensitivity	455 nm to 873 nm, 14 nm per band
Instantaneous field of view	0.86 mrad, (1.3 m at 1500 m altitude)
Number or looks, and zenith angles.	7 looks, 45 deg, forward to 45 deg. aft in 15 degree increments

Table 7.3 Sensor specification for the Advanced Solid-Sate Array Spectroradiometer (ASAS).

For sensors such as ASAS the directional measurements are all made in the plane of the orbit. In order to make measurements at different azimuth angles as well as different zenith angles, side-looking instruments are required. Thus, the practical maximum number of different directional measurements for any given target is currently probably of the order of 10, and all may be restricted to be in same plane.

7.2.2. Invertibility of SAIL with noise-free data

The invertibility of the SAIL model using several different inversion techniques was examined using only seven directional measurements. Noise-free measurements were simulated by running the model forward using a given set of parameters, to give canopy reflectance predictions. These predictions were then assumed to be the measurements, the original parameter set was "forgotten", and the model inverted in an attempt to estimate the original parameter set. This tests whether the model is invertable using the optimization technique and the given number of measurements.

Table 7.4 describes the SAIL parameters, and their possible conditions during the inversion. The geometrical parameters, θ_s, θ_o and Ψ are fixed by the measurements, the other parameters were either allowed to vary freely, constrained to lie within fixed bounds, or fixed, during the inversion.

Parameter	*Nature during inversion*
LAI (L)	Freely varying, fixed or constrained
Leaf reflectance (ρ)	Freely varying, fixed or constrained
Leaf transmittance (τ)	Freely varying, fixed or constrained
Soil reflectance (ρ_s)	Freely varying, fixed or constrained
Beta function LAD parameters (μ, η)	Freely varying, fixed or constrained
Ratio of diffuse to total incident radiation (f_{SKYL})	Freely varying, fixed or constrained
Solar zenith angle (θ_s)	Fixed by time and location of observation
Observation zenith angle (θ_o)	Fixed by measurement
Relative azimuth angle (Ψ)	Fixed by measurement
Wavelength (λ)	Fixed by measurement

Table 7.4 Parameters of the SAIL model and their condition during inversion

Four different optimization algorithms were compared, downhill simplex (DX), direction set (DS), simulated annealing (SA) and genetic algorithms (GA). For optimum results, the optimum set of parameters controlling the behaviour of each of these algorithms must be selected. Table 7.5 shows the main controlling parameters for each of the four algorithms. It can be seen from Table 7.4 and Table 7.5 that the number of possible combinations of optimization algorithm, optimization parameter set, implementation and model parameter set (constraints and nature of parameters) is very large. Using a systematic approach all of these combinations were investigated: however, a complete listing of all these results is not appropriate here. Instead, the key findings will be described and some general, salient points regarding the results will be made.

- With only seven directional measurements accurate inversion for the seven basic SAIL model parameters is likely to be very difficult.
- The reflectance predicted by the SAIL model is more sensitive to some model parameters than others. It is likely that these parameters will be more easily and accurately retrieved during the inversion.
- The interaction of LAI, leaf optical properties and soil reflectance is complex. A canopy with a small LAI and dark leaves over a bright soil will result in a similar reflectance to a canopy with large LAI and bright leaves over a dark soil. The leaf angle distribution interacts with the view angle and the LAI to make the soil more or less visible.
- The parameter f_{sky} determines the fraction of incident light that is diffuse. The larger the amount of diffuse light, the more uniform the illumination in the zenith and azimuth directions. This

reduces the effect of the hot-spot, flattens somewhat the characteristic bowl shape of the BRDF and increases its azimuthal symmetry.

Algorithm	*Control parameters*
Downhill Simplex	• Starting position • Convergence tolerance (f_{tol}) • Maximum number of iterations • Number of restarts
Direction Set	• Starting position • Convergence tolerance • Maximum number of iterations • Number of restarts
Simulated Annealing	• Starting position • Convergence tolerance (f_{tol}) • Maximum number of iterations (K) • Initial temperature (T_0) • Function evaluations at each temperature (f_{lev}) • Temperature reduction factor (α) • Implementation specific parameters
Genetic Algorithms	• Genome representation • Genetic algorithm type • Genome operators • Initial Population • Fitness scaling • Selection scheme • Mutation probability and implementation • Cross-over probability and implementation • Survival probability • Number of generations • Elitism

Annealing schedule

Table 7.5 List of controlling parameters for the four optimization algorithms used in this study.

7.2.3. Results

Figure 7.2 shows the results of downhill simplex inversion of the SAIL model with a beta function leaf angle distribution, all parameters kept free and using noise-free simulated data, as described earlier. Seven directional measurements were simulated in the near infrared (NIR) waveband. Five different parameter sets for the SAIL model were chosen in a semi-random manner. Typical ranges for each parameter were determined from literature searches and published datasets such as FIFE, HAPEX-Sahel and BOREAS. The precise values for each parameter were then chosen at random from within these ranges, to create the dataset shown in Table 7.6. The five values for each parameter

were used in combination, with the exception of the LAD parameters μ and ν, which were always used in the five *pairs* indicated. To prevent the procedure optimizing parameters to physically unrealistic values during inversion, the parameters were constrained to lie within the maximum and minimum range indicated in Table 7.7.

Parameter					
L	0.57	1.23	2.0	3.6	4.4
ρ	37	39	42	44	48
ι	53	50	49	47	45
ρ_s	10	15	20	25	30
f_{sky}	0.05	0.1	0.2	0.25	0.3
μ	2.8	1.7	1.1	3.3	0.43
ν	1.7	2.8	1.9	3.3	0.43

Table 7.6 The five values for each of the SAIL parameters used in various combinations in the inversion procedure.

As might be expected for an inversion with seven unknowns and only seven measurements, the results shown in Figure 7.2 as very poor. f_{sky} tended to be underestimated, whilst ρ_s was overestimated. There was also a tendency for ρ to be underestimated and ι to be overestimated. The inversion procedure was unable to estimate the LAD parameters correctly with both μ and ν being significantly overestimated. LAI was both underestimated and overestimated, with, in general, the error being larger than 15%. From these results, it appears that the inversion was unable to retrieve parameters accurately regardless of the dataset or initial guesses used. Careful examination of the individual inversion results shows that the optimizations tended to become trapped along the imposed constraints.

Parameter	*Min*	*Max*
L	0.25	5
ρ	30	60
ι	30	60
$\rho + \iota$	2	100
ρ_s	10	40
f_{sky}	0.05	0.6
μ	0.5	3
ν	0.5	3

Table 7.7 Constraints applied to the SAIL parameters.

The merit function values corresponding to the results of Figure 7.2 are shown in Figure 7.3. In this figure, the results from each run are plotted along the circumference, with the distance of each point from the centre increasing with increasing residual value for the merit function. The figure shows that for the vast majority of runs, very small values for the merit function were obtained, indicating a good fit between the modelled and measured data. Comparing Figure 7.2 and Figure 7.3, it is clear that the

inversion runs are repeatedly becoming trapped in local minima, but this was not unexpected with seven unknowns and only seven measurements.

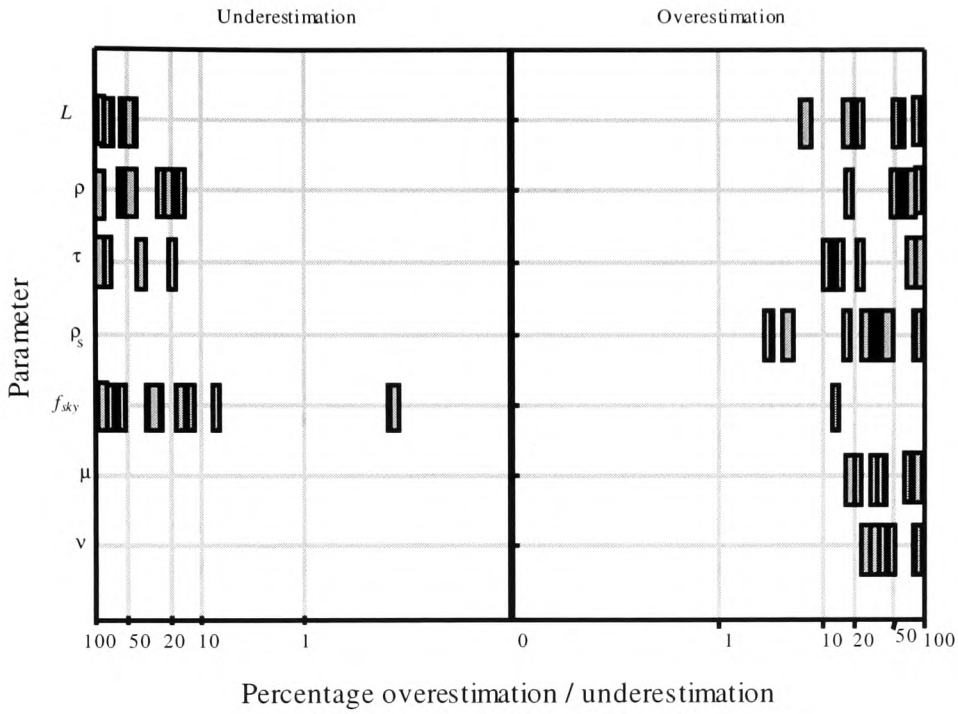


Figure 7.2 Results of downhill simplex inversion of the SAIL model, with all parameters free. The leaf angle distribution was represented by the two-parameter beta function. Noise-free *simulated* measured data was used in the NIR waveband, with seven looks, and simple constraint. Five different parameter sets were used. Each bar represents the average of five runs.

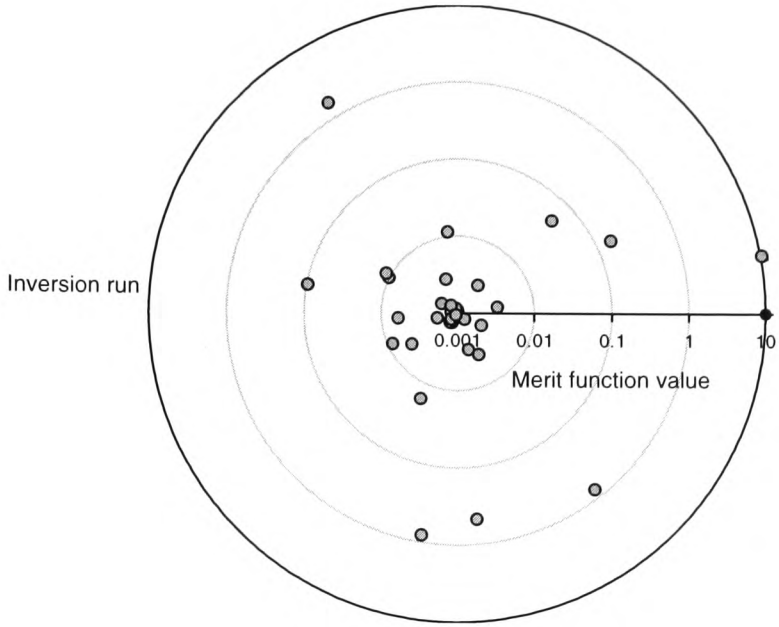


Figure 7.3 The value of the merit function after the inversion runs shown in Figure 7.2.

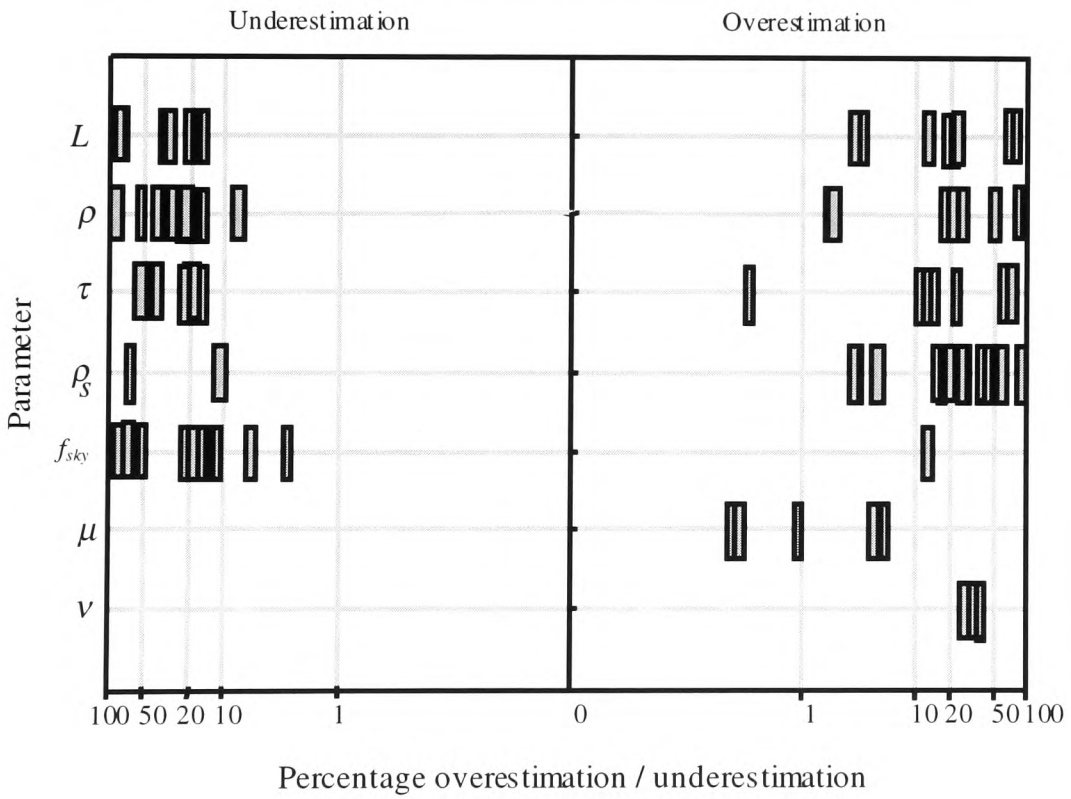


Figure 7.4 Simulated annealing inversion (SA1) of the datasets in Figure 7.2, using the annealing parameters of Table 7.13.

The performance of simulated annealing optimization (SA1) (SA1 is simulated annealing with self regulation step distribution, see Chapter 6) on the same dataset is shown in Figure 7.4. The annealing schedule parameters for these optimization runs are shown in Table 7.13. It can be seen from Figure 7.4 that although simulated annealing performed significantly better that downhill simplex, the results are still quite poor, considering that noise-free simulated data were used. With the exception of v, simulated annealing produced somewhat better results for all parameters. The majority of LAI retrievals were within 20% of the true value, with several results having errors of the about 3%. The average accuracy of retrieval of ρ and τ increased, with errors of less than 10% occurring in some cases. Soil reflectance still tended to be overestimated and f_{sky} underestimated, but the average errors were smaller than for downhill simplex. Of the LAD parameters, retrieval of μ was significantly improved with errors of less than 8%, however, retrieval of v was not improved with overestimations of 20-50% occurring in all cases. The values of the merit function for the simulated annealing runs were again very small, indicating that the algorithm was finding a local minimum, not the true global minimum.

7.3. Inversion Using Directional and Spectral Data - The PROSAIL Model

Perhaps not surprisingly, even using powerful and computationally intensive optimization techniques, such as simulated annealing, retrieval of the SAIL model parameters using only seven measurements is not possible. Clearly, more information is required in the form of measured data. Because of the limitation of remote sensing satellite design, the most fruitful source of further information is not additional directional data, but spectral data. The use of spectral information in the form of vegetation indices formed the basis for early remote sensing studies of vegetation surfaces. It is the contrast in spectral reflectance behaviour in visible and infrared wavelength between vegetation and its soil background that makes spectral information so useful. The use of traditional nadir-viewing instruments to measure vegetation indices is hindered by the signal saturation that occurs when the vegetation canopy reaches closure. Combining multi-spectral information with bidirectional reflectance measurements therefore may yield valuable extra information about vegetation canopies.

The parameters of SAIL that are wavelength dependent are the leaf reflectance, leaf transmittance and soil reflectance. The ratio of diffuse to total incident radiation f_{sky} is also dependent upon wavelength, although rather less strongly than the leaf and soil optical properties. The use of multispectral data with the SAIL model alone would not achieve an improvement in optimization performance because the inversion routine would have to calculate ρ, τ and ρ_s for each wavelength in the dataset, thus actually increasing the number of unknowns. Therefore to enable the use of spectral data, in addition to directional information, the SAIL model needs to be modified to make all parameters independent of wavelength. This can be achieved by coupling SAIL to models of leaf and soil optical properties.

Chapter 4 describes the PROSPECT model of leaf reflectance and how it may be integrated into SAIL to produce PROSAIL, a BRDF model in which all parameters except the soil reflectance are independent of wavelength. These parameters are given in Table 7.8. If the inversion of the SAIL and PROSPECT models is possible then it is likely that inversion of PROSAIL using directional and spectral data is also possible, provided that the soil reflectance is known.

	Leaf area index	L
	Leaf structure	N
PROSPECT model parameters	Leaf chlorophyll content	C_{ab}
	Leaf water content	C_w
	Leaf protein content	C_p
	Leaf cellulose and lignin content	C_c
	Mean leaf angle	$\bar{\theta}$
	Soil reflectance	$\rho_s(\lambda)$
	Diffuse radiation fraction	f_{sky}
Sun-sensor geometry parameters	Solar zenith angle	θ_s
	View zenith angle	θ_o
	Relative azimuth angle	ψ

Table 7.8 Input parameters of the PROSAIL model

Figure 7.5 shows the results of 100 optimizations using simulated annealing of the PROSPECT model using simulated data. Various random combinations of the five parameters were generated, between the maximum and minimum values given in Table 7.10. A bias was introduced to ensure that at least 25% of the combinations represented combinations of values likely to be found for dry leaves. Again, to generate the simulated data the model was run forward to generate a set of leaf reflectance and transmittance values in the spectral range from 400 nm to 2000 nm. The parameters were *forgotten* and the optimization routine was then used to attempt to find the parameters by minimizing a merit function. The results of the inversion were excellent, with the vast majority of errors being less than 0.1%. Errors of these magnitudes have a negligible effect on the modelled spectra. Sensitivity analysis of the PROSPECT model (Chapter 4) shows that it is relatively insensitive to the parameters C_c and C_p, unless the leaf water content is small. In spite of this, the optimization routine was able to produce excellent estimates of these two parameters even for moist leaves.

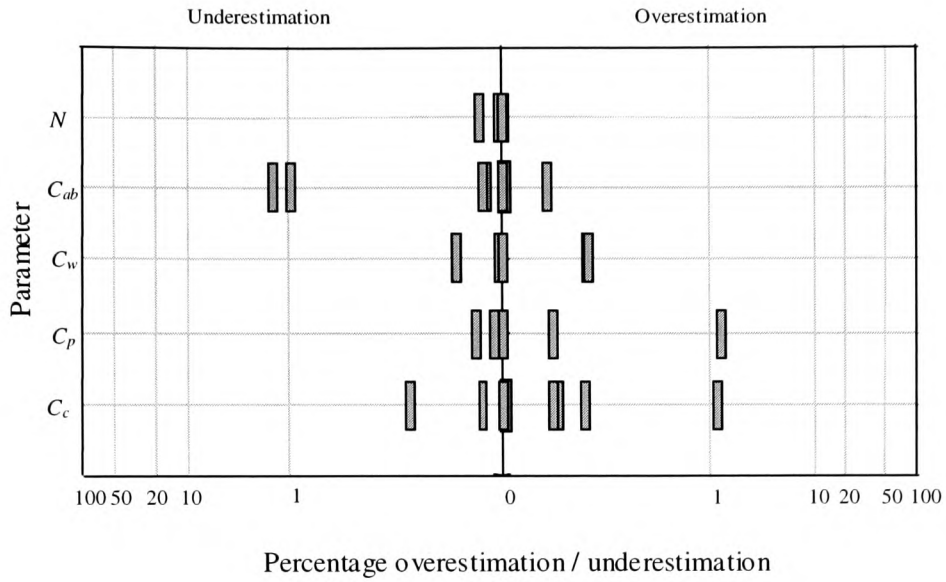


Figure 7.5 Simulated annealing (SA2) inversion of the PROSPECT-REDUX model using simulated data.

Figure 7.5 clearly indicates that PROSPECT is totally invertible using simulated data. However, any inadequacies in the representation of leaf reflectance and transmittance spectra will only be illustrated though inversion with measured data. In Chapter 4, the results of inversion of PROSPECT on measured leaf reflectance data for soybean, millet and aspen have already been presented in Figures 4.9 to 4.12, in general showing excellent results. One may, therefore, have confidence in the use of PROSPECT in inversion studies.

7.3.1. The Soilspect model

The combined SAIL and PROSPECT models may be used to predicted the multi-spectral bidirectional reflectance of vegetation provided the spectral reflectance of the soil is known. The SOILSPECT+ model described in Chapter 4 is capable of modelling the spectral reflectance of soils for a range of soil water contents, using seven parameters given in Table 4.2. Of these seven parameters, the single scattering albedo ω is a function of wavelength and the soil water layer thickness, l is a function of soil water content. The other parameters are assumed to be independent of both wavelength and soil water content. The SAIL model only requires Lambertian soil reflectance so the four bidirectional scattering parameters b,b',c,c' are fixed, leaving three parameters varying. The number of datasets of the spectral response of soil reflectance to water content is very limited indeed making accurate simulations difficult. Moreover, theory suggests that the importance of soil reflectance in BRDF inversion (Chapter 3) can be large, since, for example, the reflectance of bright leaves over a dark

substrate for a small LAI approximates the reflectance of dark leaves over a light subject for a large LAI.

In this study soil reflectance was not calculated by the combined PROSAIL-SOILSPECT+ model (known as PROSAIL+) during inversion. Instead, the SOILSPECT+ model was used to pre-compute the spectral reflectance of a number of soil types to form a look-up table. The look-up table consists of fixed combinations of the three parameters ω, h and l and the associated spectral reflectance. The range and resolution for ω and hence for the soil reflectance, was 300 nm to 2300 nm in 20 nm steps. In common with many machines, Sun workstations store single precision numbers as one 32-bit word. Each complete row of the look-up table therefore occupies 100 bytes. Ten different soil spectra were chosen with five roughness values and five water contents, making 250 complete soil spectra in all. The look-up table is therefore easily held in memory on most computers. The look-up table parameters were selected from values in the literature, by simulated annealing inversion from the HAPEX-Sahel, OTTER and BOREAS datasets, or from the NASA World Wide Web archive.

The fraction of diffuse radiation was calculated from the horizontal visibility, V, using a another look-up table. The values in the look-up table were calculated using 5S model (Tanré *et al.* 1990 – Chapter 4). Each row of the table contained the value of the f_{sky} in 50 nm intervals from 300 nm to 1400 nm for a given value of the visibility, V. If the value of f_{sky} was known from measurement at one wavelength, or if the wavelength independent parameter, V, was known, then the table could be used to infer its value at other wavelengths. If no information about the visibility or fraction of diffuse radiation was available, then the best fit row of the table was determined by the optimization algorithm.

7.3.2. Inversion of PROSAIL+ using noise-free data

The invertability of the combined PROSPECT–SAIL-SOILSPECT+ model (known as PROSAIL+) was tested by attempting to optimize the model using noise-free simulated data. Fifty test datasets were generated in the same manner as before by forward running of the model with a known parameter set, selected at random from combinations of parameters indicated in Table 7.6, Table 7.7 and Table 7.10, to provide the "*measured*" dataset. The *measured* data consisted of seven directional measurements in each of six wavebands. The wavebands were arbitrarily chosen to mimic those of Landsat TM, and are listed in Table 7.9.

Waveband	*Min (nm)*	*Max (nm)*
1	450	520
2	520	600
3	630	690
4	760	900
5	1400	1700
6	2000	2300

Table 7.9 The wavebands used in the optimization runs.

The values used for the parameters LAI, μ and v and the limits imposed on them were the same as in Table 7.6 and Table 7.7, respectively. PROSAIL contains 5 parameters from PROSPECT used to control the leaf optical properties; the set of values for these parameters used in the inversion are shown in Table 7.10, together with the maximum and minimum values of each parameter allowed during the inversion. The application of maximum and minimum constraints was applied in a slightly different manner to the previous inversion attempts. When the algorithm optimized a parameter to a value outside its constraints, the merit function was given a very large value proportional to the distance of the parameter value beyond its constraint. Thus, the algorithm was directed back into the "allowed" regions of the feature space, and larger transgressions were penalized more than smaller transgressions.

Parameter						*Min*	*Max*
N	1.0	1.3	1.7	2.1	2.5	1.0	3.0
C_{ab} [μg cm^{-2}]	15.0	25.0	32.0	41.0	50.0	5.0	75.0
C_w [cm]	0.009	0.017	0.024	0.032	0.039	0.009	0.0400
C_p [g cm^{-2}]	0.0006	0.0008	0.0011	0.0014	0.0015	0.0005	0.0017
C_c [g cm^{-2}]	0.0019	0.0025	0.0031	0.0037	0.0042	0.0017	0.0044

Table 7.10 The five values for each of the parameters in PROSAIL that control leaf optical properties, and the maximum and minimum value of each parameter that were allowed in the optimization runs.

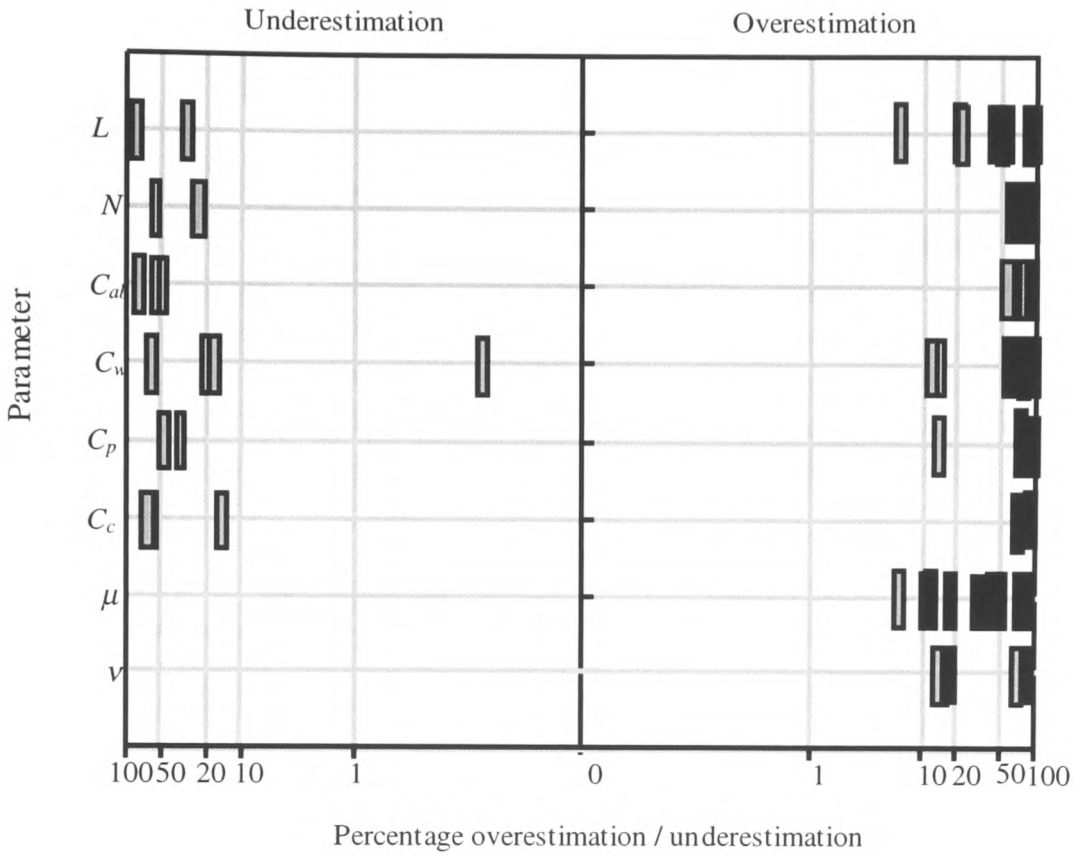


Figure 7.6 Downhill simplex optimization of the PROSAIL model, using simulated data.

7.3.2.1. Downhill simplex method

Figure 7.6 shows the results of downhill simplex optimizations of the test datasets. Of the 50 tests datasets, only 25 achieved convergence. The estimation of parameters for those datasets that did converge was poor, with results, in general being worse than for the SAIL model alone (Figure 7.2). There was a marked propensity to overestimate all parameters, which is indicative of instability in the inversion process. Detailed analysis showed that, regardless of the initial guess for the parameter set, the algorithm had a tendency to increase parameter values until the upper bound was reached. This made the use of repeated restarts from new positions ineffective. The values of the merit function for those datasets that did converge were slightly worse than for the earlier study with the SAIL model alone and only seven directional measurements. This is discouraging since, although the combined PROSAIL model is significantly more complex than that used in the previous study, the information content of the measurements is larger. This behaviour suggested a possible bug in the code of the program, but attempts to optimize a series of simple mathematical formulae were successful. Therefore, one must conclude that for this problem, the downhill simplex algorithm is unable to find

good minima, either because the problem is too difficult (i.e. too many local minima exist) or because this technique is unsuited to this particular problem.

7.3.2.2. Direction set method

Using the direction set optimization routine on the same test datasets produced somewhat better results, as shown in Figure 7.7. although the results are disappointing, considering the use of noise-free simulated data. Convergence was achieved for all datasets, but there was still a tendency to overestimate parameters. Retrieval of LAI was good with errors for the most part in the range 2% to 24%. The parameters, C_c and C_p, were rather poorly estimated, but considering the lack of sensitivity of the PROSPECT model to these parameters, unless the leaves are relatively dry, this is not surprising.

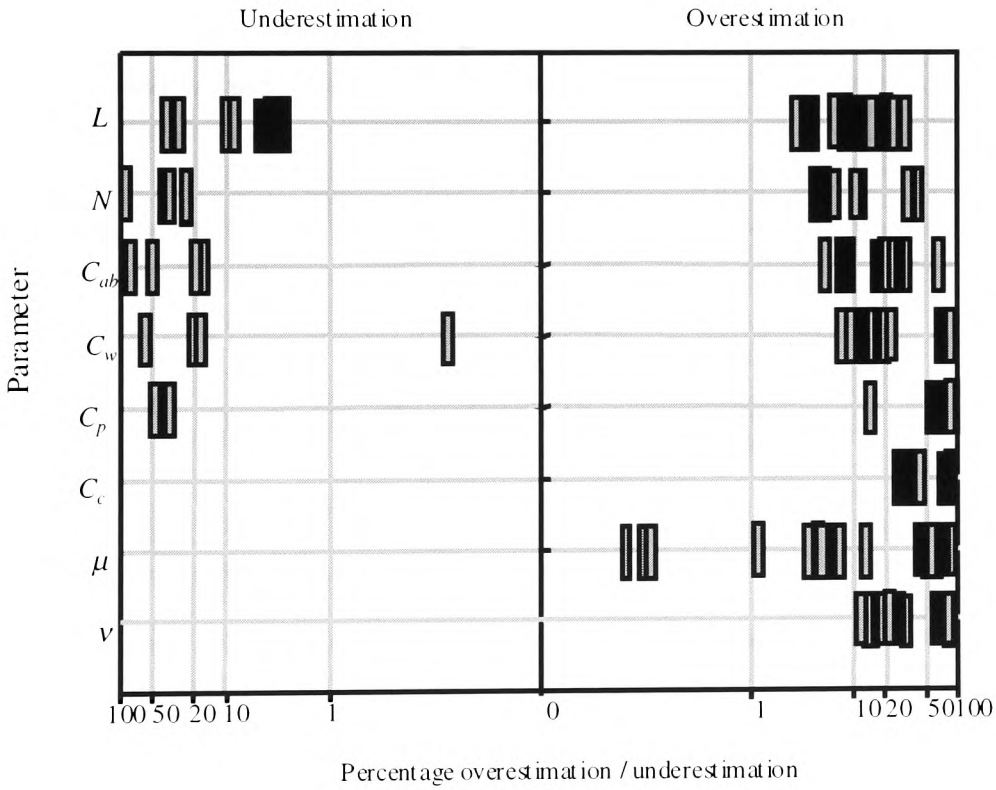


Figure 7.7 Direction set optimization of the PROSAIL model: max iterations = 50, number of restarts = 5, $f_{tol} = 10^{-5}$

The distribution of retrieved values is skewed with the majority of values being overestimated but with a scattering of underestimations for LAI, N and C_{ab} suggesting that the algorithm was not searching the whole of the feature space.

The values for the leaf angle distribution parameters, μ and ν were poorly retrieved. In several instances one or other (usually μ) of the parameters was found to within 10% accuracy, but the estimated value for the other parameter had an error of larger than 25%. This can lead to large errors in the predicted leaf angle distribution, which is a combination of these two parameters (Table 7.11).

Distribution	*Ave angle*	*μ*	*ν*
Planophile	26.76°	2.77	1.17
Erectophile	63.24°	1.17	2.77
Plagiophile	45.00°	3.33	3.33
Extremophile	45.00°	0.43	0.43
Uniform	45.00°	1.00	1.00
Spherical	57.30°	1.10	1.93

Table 7.11 Values of the beta distribution parameters μ and ν, and the average leaf angle for six different leaf angle distributions.

The beta distribution for leaf angle (Goel and Strebel, 1984) represents the distribution of leaf angle very well for many canopies. Unfortunately, the two parameters are not completely independent (Equation 7.3), and the function $f(\mu,\nu,\theta)$ does not vary smoothly with μ and ν. Neither does the behaviour of any one of the two parameters alone vary in proportion to the shape of the leaf angle distribution, i.e. increasing μ does not correspond to an increasingly erectophile distribution, or vice versa. Instead, it is the combination of the two values, μ and ν that determines the distribution: the values of μ and ν for the six *standard* leaf angle distributions are given in Table 7.11.

$$f(\mu,\nu,\theta)=\frac{1}{360\times 90}\frac{\Gamma(\mu+\nu)}{\Gamma(\mu)\Gamma(\nu)}\left[1-\frac{\theta}{90}\right]^{\mu-1}\left[\frac{\theta}{90}\right]^{\nu-1} \tag{7.3}$$

The lack of independence and smooth variation of μ and ν makes it difficult for an inversion algorithm to follow the function toward a minimum.

The simpler, elliptical leaf angle distribution (Campbell, 1986, 1990; Wang and Jarvis, 1988), requires only one parameter (the mean leaf angle, $\bar{\theta}$) and considers the distribution of leaf area to be like the surface of a prolate or oblate spheroid, (Equation 7.4).

$$f(\theta)=\frac{2\chi^2\sin\theta}{\Lambda\left(\cos^2\theta+\chi^2\sin^2\theta\right)^2} \tag{7.4}$$

Where: $\Lambda\approx\chi+1.774(\chi+1.182)^{-0.733}$ and $\bar{\theta}\approx 9.65(3+\chi)^{-1.65}$

Whilst the elliptical distribution is unable to model distributions that cannot be approximated in this manner, (e.g. plagiophile), or very complex distributions, where the leaf angle distribution exhibits two or more clusters, it is able, to represent the LAD of many plant species reasonably well (Norman

and Campbell, 1989, and Figure 7.8). Both the beta and elliptical distribution functions are less peaked than the measured distribution, and although the beta function fits the data rather better than the elliptical distribution, the fit of the elliptical function is within the typical measurement error for LAD. The elliptical distribution not only reduces the number of parameters to be found through inversion by one, but the relatively smooth behaviour of the function is likely to improve inversion efficiency. To test this assumption the optimizations described above were repeated replacing the beta leaf angle distribution function in SAIL with the elliptical function. The five values for $\overline{\theta}$ used are shown in Table 7.12 and the results of the optimization for three different values representing open, intermediate and closed canopies, are shown in Figure 7.9.

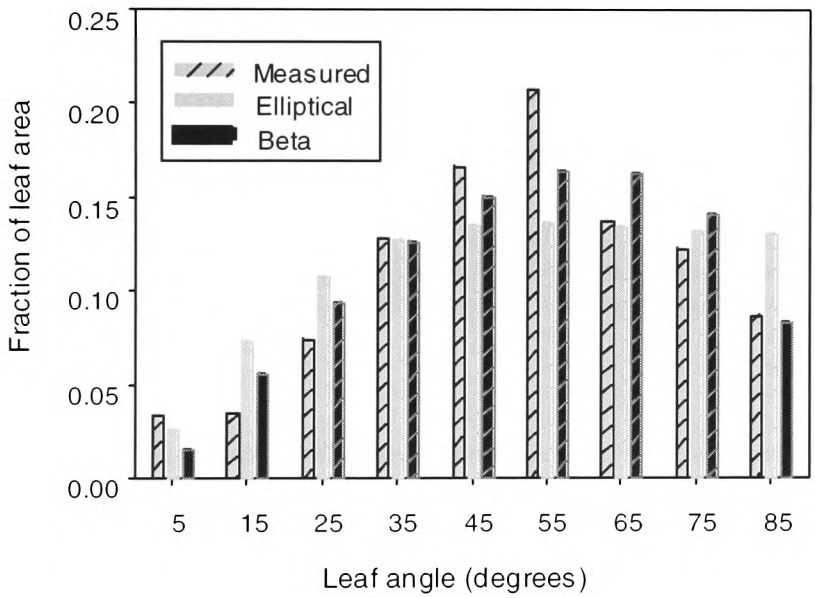


Figure 7.8 Comparison of the leaf angle distribution for soybeans measured by Ranson *et al.* (1985) (28 August 1980) and the values predicted by the elliptical and beta function distribution models using parameters calculated from the measured data.

Parameter						*Min*	*Max*
$\overline{\theta}$	26.76°	35.0°	45.0°	57.30°	63.24°	20.0°	70.0°

Table 7.12 The five values for the mean leaf angle, $\overline{\theta}$ in the inversion studies using the elliptical leaf angle distribution function.

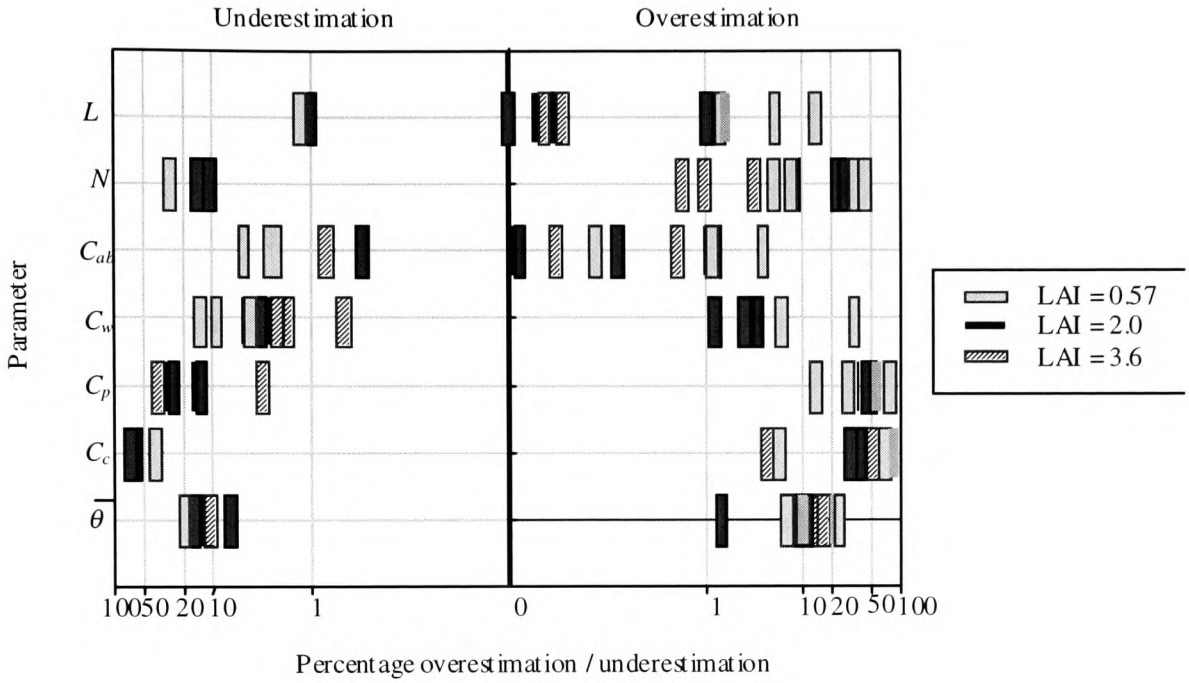


Figure 7.9 Results of direction set optimization of the noise free simulated dataset, using the elliptical leaf angle distribution function, for three values of LAI and random combinations of other parameters.

Figure 7.9 shows that the retrieval of $\overline{\theta}$ is significantly better than for the beta distribution parameters, μ and ν, with errors typically being less than 20%. Moreover, there is also an improvement in the retrieval of the other parameters, indicating that the complex response of LAD to μ and ν was hindering the optimization algorithm. Once again, the accuracy of determination of C_p and C_c is relatively poor; however, since for all but dry leaves, the leaf optical properties are only weakly sensitive to these parameters, somewhat larger errors on these parameters may be expected.

With the exception of C_p and C_c the determination errors for all parameters decreased with increasing LAI. For small values of LAI, the relative contribution of soil to the overall reflectance becomes much larger increasing the difficulty of parameter retrieval. For crops, small values of LAI frequently correspond to plants at an early growth stage, when the canopy is open and large areas of the soil are visible between individual plants. The SAIL model is known to work less well for small leaf area, open, canopies because the underlying assumption of horizontal homogeneity is violated.

Modelled reflectance is only sensitive to C_w for wavelengths longer that 1200 nm, with maximum sensitivity occurring around 1600 nm (Chapter 4). Therefore, only two of the test dataset wavebands can be expected to show sensitivity to this parameter. In spite of this, retrieval results for this

parameter were quite good, typically in the range 5 to 10% for the larger two LAIs. Sensitivity of the model to C_{ab} is restricted to the range 500 to 750 nm, again corresponding to only two wavebands. However, sensitivity of the model to C_{ab} is quite high (Chapter 4) and this is reflected in the excellent retrievals obtained.

Retrieval of LAI was very good, especially for the two larger LAIs, errors being less than 1.2% in all cases. However, it must be remembered that simulated data were used: for *real* measurements, both signal noise and inaccuracies in the model make inversion much more difficult. To have confidence in the effectiveness of an inversion technique with *real* data, the technique should be able to retrieve parameters with simulated data and no noise to extreme accuracy.

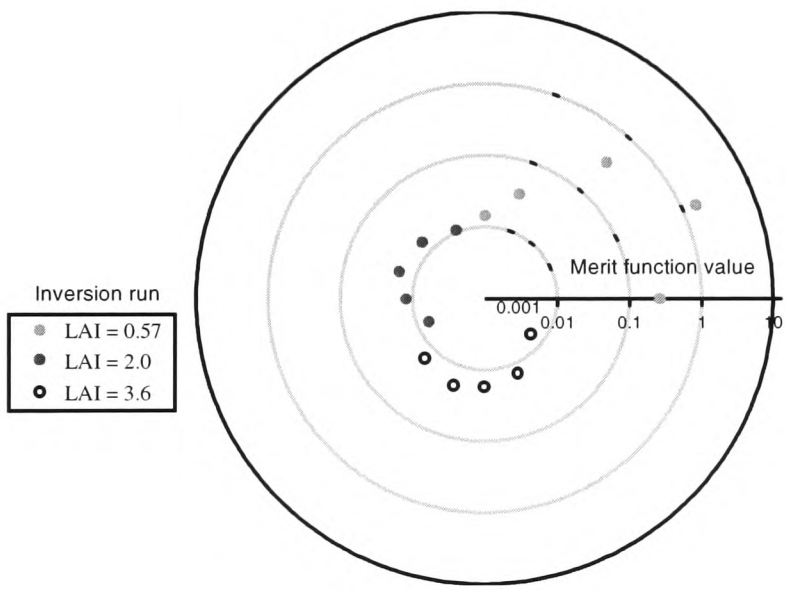


Figure 7.10 Values of the merit function for the direction set inversion of the noise-free simulated dataset, for the three LAI values of Figure 7.9.

Figure 7.10 shows the values of the merit function corresponding to the parameter retrievals of Figure 7.9. The performance of the DS algorithm is clearly worst for the small LAI case, whereas the performance at the two larger LAI values is about the same. Although the values for the merit function for the larger LAI values are quite small when using noise-free, simulated data, negligible values for the merit function might be expected.

7.3.2.3. Simulated annealing method

The optimizations shown in Figure 7.9 were repeated with identical datasets using simulated annealing. Using the default set of schedule parameters shown in Table 7.13, excellent results were

obtained for both SA1 and SA2 (SA2 is simulated annealing with downhill simplex search, see Chapter 6) with all parameters being perfectly retrieved (errors on all parameters being less than 0.0001%). Results such as these do not make for very interesting graphs or tables, so these are not shown here. It is suffice to say that the PROSAIL model is clearly totally invertable given the measurement conditions of this synthetic experiment. However, all real-world measurements contain noise. To test the invertability of the model under noisy conditions they were repeated using simulated noisy data.

SA1		*SA2*	
Parameter	*Value*	*Parameter*	*Value*
K	10000	K	10000
T_0	1000	T_0	1000
M	120	f_{tol}	1e-15
χ_t	0.70	f_{lev}	200
χ_s	3.0	α	1.0

Table 7.13 Annealing schedule parameters.

7.3.3. Inversion of PROSAIL+ Using Noisy Data

To generate the noisy data, the same datasets as before were used to parameterize the PROSAIL+ model for forward running to produce *measured* reflectances. To generate the noisy measured data, 5% Gaussian-distributed, random noise was added with a 0.5% systematic bias. This was achieved by skewing the distribution of the random number generator.

SA1		*SA2-α (alpha parameter cooling schedule)*		*SA2-log (logarithmic cooling schedule)*	
Parameter	*Value*	*Parameter*	*Value*	*Parameter*	*Value*
K	10000	K	10000	K	10000
T_0	5000	T_0	5000	P	0.3
M	200	f_{tol}	1e-15	W	3.5
χ_t	0.70	f_{lev}	200	C	2.0
χ_s	3.0	α	1.0		

Table 7.14 First guess annealing schedule parameters used for the noisy dataset

7.3.3.1. Simulated annealing

As described in Chapter 6, the choice of annealing schedule parameters can have an important effect on the performance of simulated annealing optimizations. The first guess schedule parameters for SA1 and SA2 are given in Table 7.14, and the results of the optimization runs using these parameters are given in Figure 7.13, and Table 7.15. The parameters in Table 7.14 were derived from experience gained in studying the simulated annealing technique on a variety of problems during this study (e.g.

see Chapter 6). Most of the retrievals have errors of 20% or larger. The smallest individual errors occur for SA2-log, but the results for this algorithm also exhibit a larger degree of scatter than for SA1 and SA2-α. This is reflected in the average retrieval errors given in Table 7.15. The errors for LAI, C_{ab} and C_w are significantly lower for SA1 – these may be regarded as the most important parameters so far as vegetation modelling is concerned. The error for average leaf angle $\overline{\theta}$ is rather lower for SA2-log, than for SA2-α and is also slightly lower than SA1. The errors for N, C_p and C_c are large but this is not unexpected considering the low sensitivity of the PROSPECT model to these parameters. The retrieval error for LAI using SA2 is disappointing, but careful examination of the results shows that a significant component of the average error is a result of poor performance by the algorithm for small values of LAI in the test dataset.

The results in Table 7.15 are discouraging, but can improvements be made by tuning the annealing schedule parameters? Extra care must be taken when choosing the schedule parameters for SA1, since an injudicious choice of K, M and χ_t can results in the temperature failing to approach zero before the maximum number of iterations is reached. This is illustrated in Figure 7.11, which shows the rate of cooling for three different values of χ_t. With, $M = 105$, $T_0 = 5000$, $\chi_s = 3.0$ the temperature is still larger than 1900 when the maximum number of iterations is reached. For values of χ_s smaller than 0.9 the rate of cooling is very rapid which risks quenching the system. As described in Chapter 6, M must be chosen large enough to allow sufficient statistics for the random walk. By trial and error $M = 15n$, where n is the number of dimensions (unknowns) of the system, appears to be a good benchmark figure (Figure 7.12).

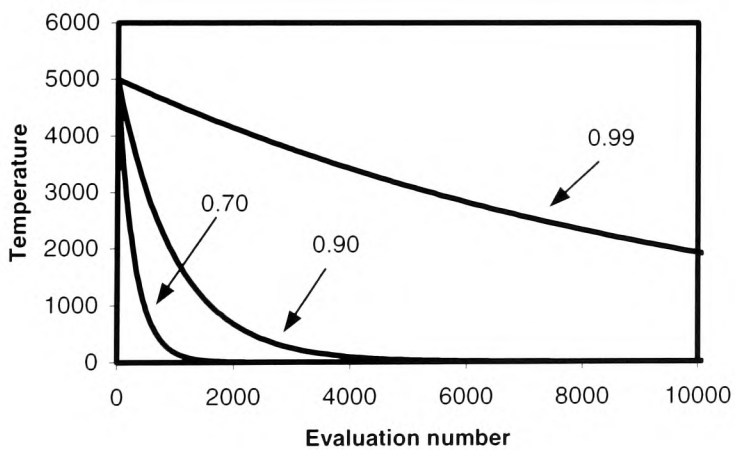


Figure 7.11 The cooling schedule for SA1 For three different values of χ_t (0.70, 0.90 and 0.99) with $K = 10000$, $M = 105$, $T_0 = 5000$, $\chi_s = 3.0$.

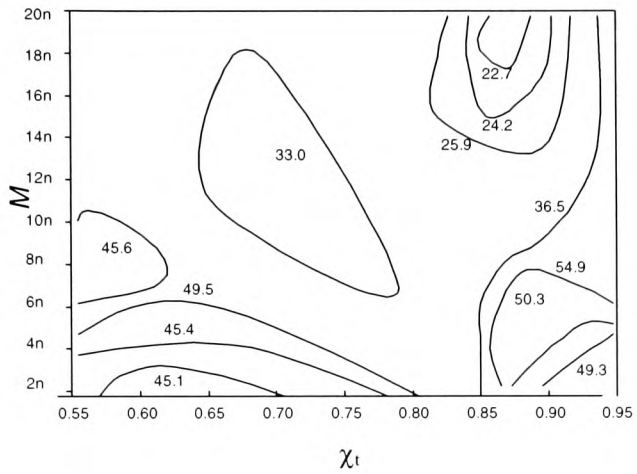


Figure 7.12 Effect of changes in two schedule parameters on the SA1 optimizations of the PROSAIL model using the noisy dataset ($n = 7$, is the dimensionality of the system).

Figure 7.12 shows the contour map of the variation of the merit function for the noisy dataset with the two SA1 schedule parameters, χ_t and M. For values of $\chi_t < 0.7$ changing M has little effect. This is probably because the rapid initial cooling inherent with smaller values of χ_t, quenches the system into the region of a local minimum. Similarly, for small values of M the step size of the random walk is insufficient to obtain good statistics (Chapter 6, Equation 6.30). This results in only a subset of the feature space being searched reducing the probability of finding the global minimum. Figure 7.12 seems to indicate that $M = 17n$ and $\chi_t = 0.9$ would be a good choice. Using the *specific heat* approach (Equation 6.35c) indicated that the initial choice of T_0 was far too high and that a value of around 100 was sufficient to ensure that the system was molten initially. These parameters (Table 7.16) improved accuracy of the retrievals significantly (Table 7.17).

Parameter	Average retrieval error (%) SA1	SA2(alpha)	SA2 (log)
LAI	27.38	33.91	57.03
N	43.86	44.72	46.55
cab	24.22	30.74	61.48
cw	30.19	39.08	60.60
cp	48.72	49.87	47.32
cc	47.71	53.89	88.51
$\overline{\theta}$	19.23	27.12	18.30

Table 7.15 Average retrieval error for SA1 and SA2 optimizations of the noisy dataset using the first guess annealing schedule parameters given in Table 7.14, expressed as a percentage of the parameter value.

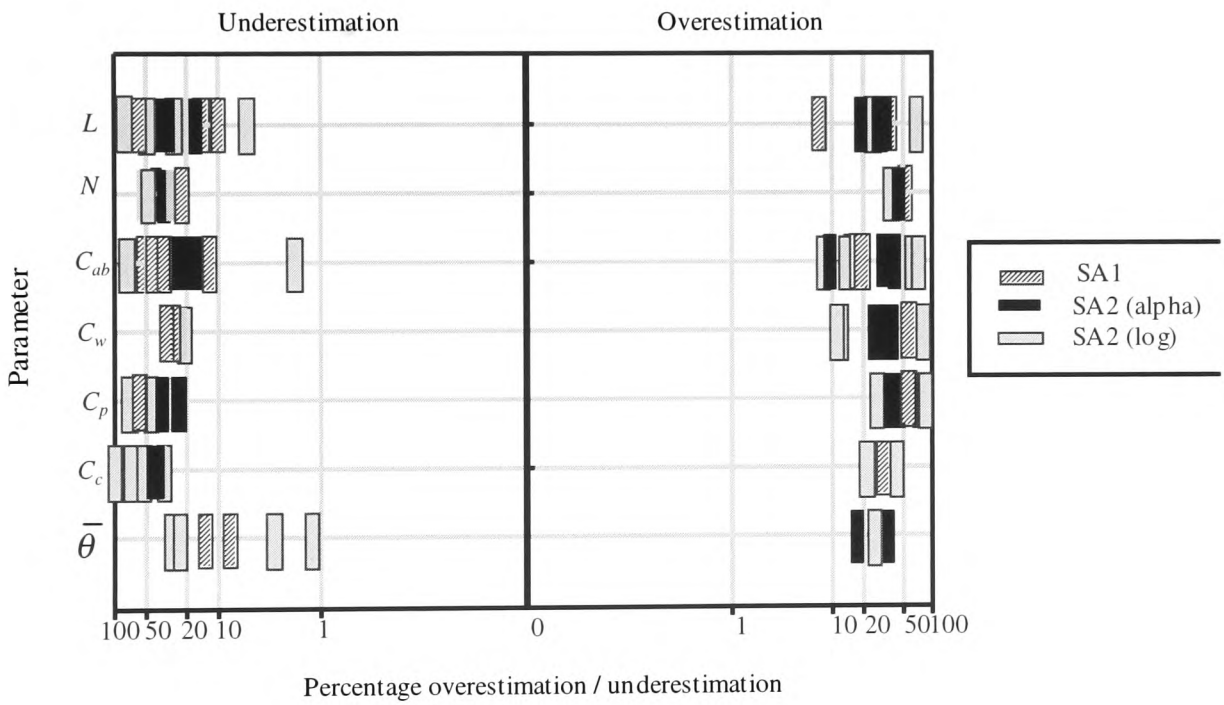


Figure 7.13 Results of simulated annealing optimizations of the noisy dataset using the first guess schedule parameter given in Table 7.14.

Figure 7.14 shows the merit function trace for SA2 for one of the combinations for the noisy test dataset, using the alpha annealing schedule given by Equation 6.35a with the first guess schedule parameters. The merit function can be seen to veer widely between high and low values, especially at higher temperatures and failed to stabilize as the temperature was reduced. The distinctive molten, freezing and solid stages are absent. This indicates that the initial temperature was too high and/or that there were too few at the lower temperatures.

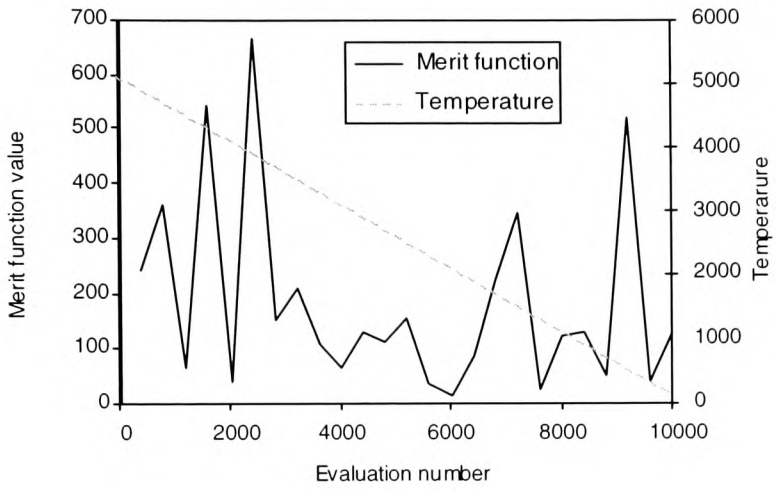


Figure 7.14 Merit function trace for one combination of the noisy test dataset using SA2 and the "alpha" schedule with the first guess schedule parameters.

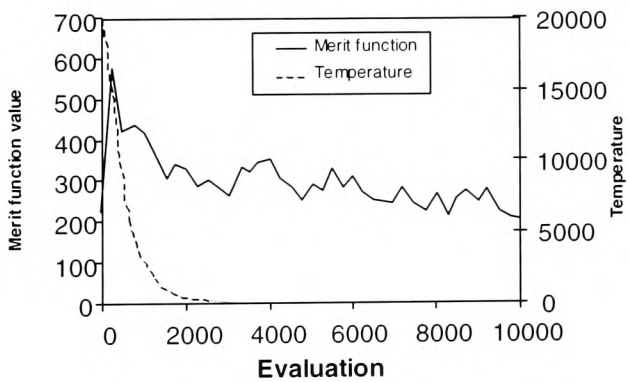


Figure 7.15 Merit function trace for one combination of the noisy test dataset using SA2 and the logarithmic schedule with the first guess schedule parameters.

The corresponding merit function trace for SA2 using the logarithmic schedule is shown in Figure 7.15. Using the logarithmic schedule and the first guess schedule parameters, the initial temperature is relatively high, but is cooled rapidly to very low values. The behaviour of the merit function suggests, once again, that the initial temperature was too high and that the rapid cooling quenched the system into the region of a local minimum, allowing only minimal further improvement over many thousands of iterations.

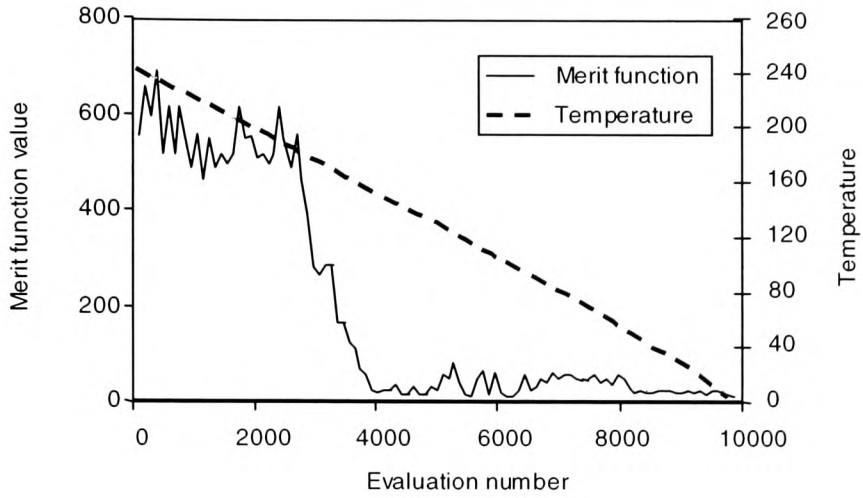


Figure 7.16 Merit function trace with near-optimum set of annealing parameters, exhibiting molten, crystallizing and solid phases.

The logarithmic schedule does not seem well suited to this problem since a gradual reduction of temperature appears to be necessary to allow the annealing algorithm to weave its way through the noisy data. The "alpha" function is better suited in this regard, and the effect on the merit function trace with the initial temperature reduced to 250 and $\alpha = 1$ is shown in Figure 7.16. This combination of schedule parameters appears to be near optimum for this problem. The molten, freezing and frozen states are clearly visible in the figure, yet the initial temperature is not so high as to *waste* iterations.

SA1		*SA2 (alpha)*	
Parameter	*Value*	*Parameter*	*Value*
K	10000	K	10000
T_0	100	T_0	250
M	120	f_{tol}	1e-15
χ_t	0.91	f_{lev}	200
χ_s	3.0	α	1.0

Table 7.16 Best fit annealing schedule parameters for SA1 and SA2-α.

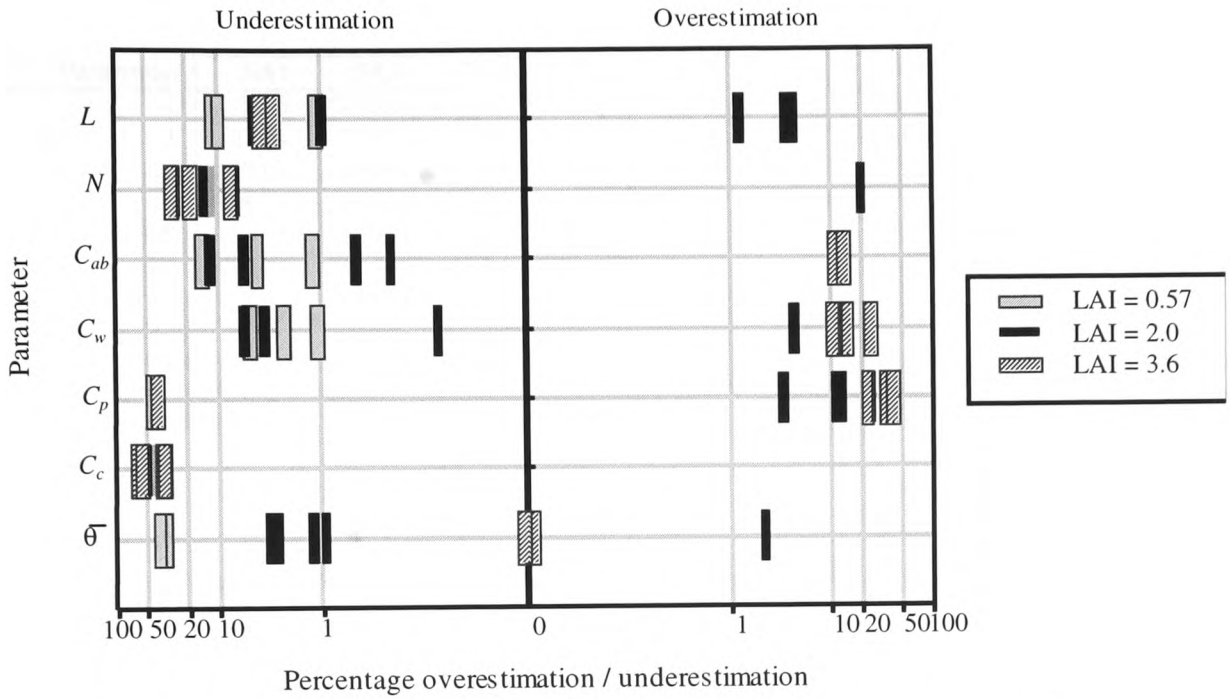


Figure 7.17 Results of SA1 optimization of the noisy dataset using the tuned schedule parameters.

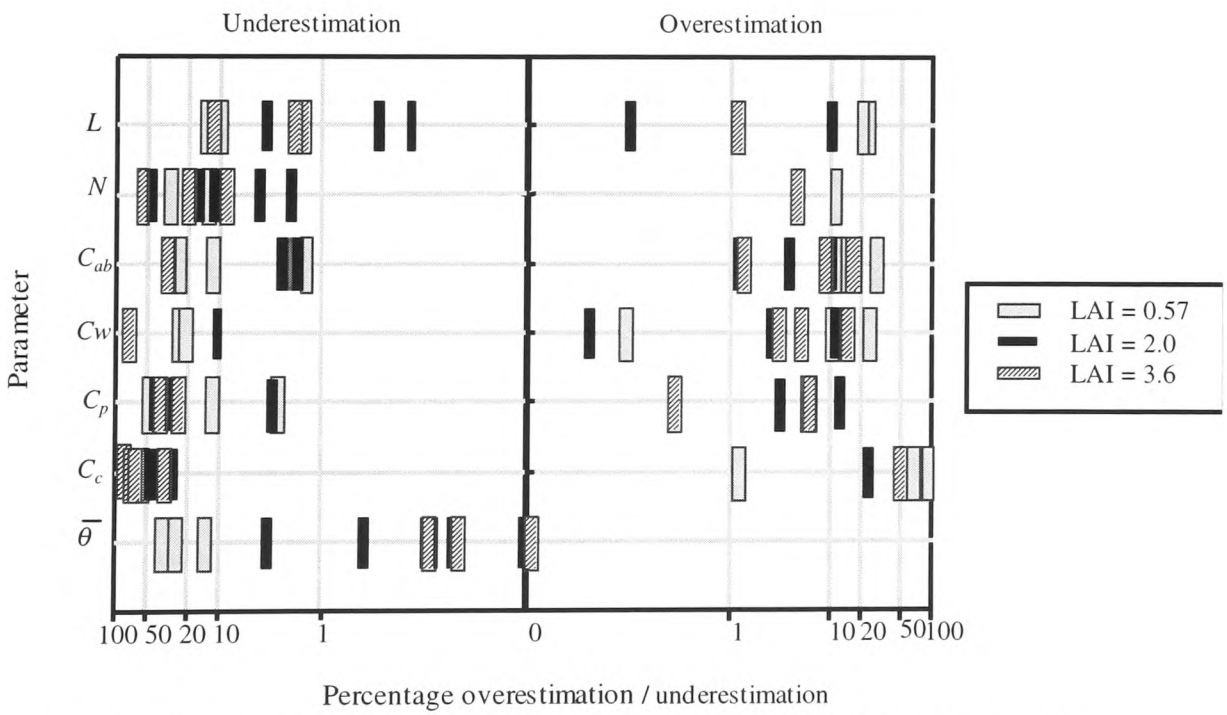


Figure 7.18 Results of SA2 optimization of the noisy dataset using the tuned schedule parameters.

	Average retrieval error (% of parameter value)							
	L = 0.57		L = 2.0		L = 3.6		Average all	
Parameter	SA1	SA2	SA1	SA2	SA1	SA2	SA1	SA2
L	5.85	15.83	3.97	3.68	5.25	11.71	5.03	10.41
N	27.64	21.91	20.93	37.88	51.74	46.34	33.43	35.37
C_{ab}	11.09	20.29	3.67	4.03	13.08	28.97	9.28	17.76
C_w	6.73	19.37	7.58	5.75	12.42	24.88	8.91	16.67
C_p	48.76	20.79	21.77	22.41	35.16	31.81	35.23	25.00
C_c	56.26	57.64	38.34	54.42	54.66	55.04	49.75	55.70
$\bar{\theta}$	42.38	34.54	1.26	2.86	0.01	0.23	14.55	12.54

Table 7.17 Average retrieval error for SA1 and SA2-α optimization of the noisy dataset using the best fit annealing schedule parameters found by trial and error, expressed as percentage of parameter value.

The improvement that can be made to the retrieval accuracy of SA1 and SA2 by adopting the "alpha" schedule, and the best fit schedule parameters given in Table 7.16, are shown in Figure 7.17, Figure 7.18 and Table 7.17. There is a marked improvement in the performance of both algorithms, although, again, the errors associated with SA1 are smaller than SA2. In particular the errors for LAI reduced from 27.38, 33.91 to 5.03, 10.41% for SA1 and SA2, respectively. SA2 exhibits the largest variation in retrieval errors, but achieves the smallest individual errors for most of the parameters. This suggests that the algorithm is slightly unstable and may benefit from some customisation for this particular problem. The best performance occurs for LAI = 2: for smaller LAIs incomplete canopy cover has the result that the relative contribution of the soil reflectance is high, for larger LAIs the problem of signal saturation become apparent (Chapter 3). The only contradiction to this is the accuracy of retrieval of mean leaf angle, which appears to increase with increasing LAI. As before, the errors for N, C_p and C_c are high because of the low sensitivity of the model to these parameters. The retrieval accuracy for LAI using SA1 is excellent throughout and the error for LAI is less than the random noise for LAI = 2. Even errors as large as 15% approach the likely field measurement error for this parameter.

7.3.3.2. Genetic algorithms

Genetic algorithms (GAs) are a very powerful but very computationally intensive optimization technique, reviewed in Chapter 6. There are a wide number of variations that may be introduced into the implementation of GAs; these are listed in Table 7.5. Parameters such as the initial population, number of generations, mutation probability and the cross-over probability are normally specified at run-time. Other parameters such as the genome operators and selection scheme are implementation dependent.

The GA software written for this study was designed to be as flexible and extensible as possible. Written in C++, using object-oriented programming techniques, it contains several built-in GA types but it is easy for users to define their own customized applications. The whole of the implementation may be customized, from the genome representation, to the selection criteria and replacement schemes. The default specific options are as follows:

Genome representation

- Real-number genome
- Binary string genome
- String genome

A real-number genome representation was chosen for all the GA optimizations in this study. This seems a natural choice for the optimization of numerical models. This representation specifies an array of bounded real number parameters, where each gene is a real number representing one of the PROSAIL model parameters.

Algorithm Type

- Simple GA (non-overlapping populations) – each generation the algorithm creates an entirely new population, of individuals by selecting from the previous population then mating to produce new offspring for the new population. Elitism is optional.
- Steady state GA (overlapping populations) – each generation the algorithm creates a temporary population adds these to the previous population then removes the worst individuals to return the population to its original size (note that this means that the new offspring may or may not survive into the next generation). The percentage of population replaced each generation is specified.

Scaling scheme

- No scaling
- Linear scaling
- Sigma truncated scaling

The scaling scheme determines how the fitness scores are derived from the merit function value. Scaling is used to keep appropriate levels of competition amongst individuals throughout a simulation; this is very important for small population sizes. At the start of GA runs it is common to have a few very high performing individuals in the population. Without scaling there is a tendency for these individuals to dominate the selection process. In this case, the merit functions values of these individuals must be scaled back to prevent them dominating. Conversely, near the end of the run the population has largely converged (i.e. most of the individuals are clustered around the optimum) and competition amongst population members is weak and the simulation tends to wander. The population average fitness may be close to the best fitness which results in average performing members getting the same number of copies in future generations as the best performers. In this case,

the merit function value of the best individuals must be scaled up to continue to reward the highest performers.

An obvious choice for a scaling function would be a linear scaling of the form:

$$f' = af + b \tag{7.5}$$

Where: f' is the scaled fitness

f is the merit function value

a and b are scaling coefficients

The coefficients a and b are chosen so that the average scaled fitness is equal to the average merit function value. This ensures that each average population member contributes one expected member into the next generation (Goldberg and Deb 1991). It is also useful to control the number of offspring given to the population member with the best merit function value, through fitness scaling. This is done as follows:

$$f'_{\max} = C_m f_{avg} \tag{7.6}$$

Where: $f'_{\max}$ is the scaled maximum fitness value

C_m is the number of expected copies for the best population member

f_{avg} is the average merit function value.

The scaling coefficients a and b are then specified in the following way.

If $$u_{\min} > \frac{C_m u_{avg} - u_{\max}}{C_m - 1}$$

then

$$a = \frac{(C_m - 1)u_{avg}}{u_{\max} - u_{avg}} \qquad b = u_{avg} \frac{(u_{\max} - C_m)u_{avg}}{u_{\max} - u_{avg}}$$

else

$$a = \frac{u_{avg}}{u_{avg} - u_{\min}} \qquad b = \frac{-u_{\min} u_{avg}}{u_{avg} - u_{\min}} \tag{7.7}$$

Where: u_{max} is the maximum value of the merit function

u_{min} is the minimum value of the merit function

u_{avg} is the average value of the merit function

For small population sizes ($50 \leq n \leq 100$), C_m is frequently chosen to be in the range 1.2 to 2.0. Figure 7.19 shows, diagrammatically, linear fitness scaling under normal conditions. Looking at the figure there do not appear to be any difficulties with linear scaling, but Figure 7.20 illustrates a potential problem. The situation shown in Figure 7.20 typically occurs near the end of an optimization run, when the average and maximum merit function values are relatively close, and when it is common to have a few quite poor individuals (known as *lethals*) in the population. The steep slope of the scaling function necessary to scale f_{max} to $2f_{avg}$ causes low fitness values to become negative. Negative fitness values are not allowed. Several ways to circumvent this problem have been proposed, the simplest of which is to reset all negative fitness values to zero. However a better technique, known as sigma truncation, uses population variance information to pre-process merit function values prior to scaling as follows:

$$f' = f - \left(\bar{f} - c\sigma\right) \tag{7.8}$$

Where: c is a constant chosen to be a reasonable multiple of the population standard deviation, (usually between 1 and 3).

σ is the population standard deviation.

$\bar{f}$ is the population average value of the merit function.

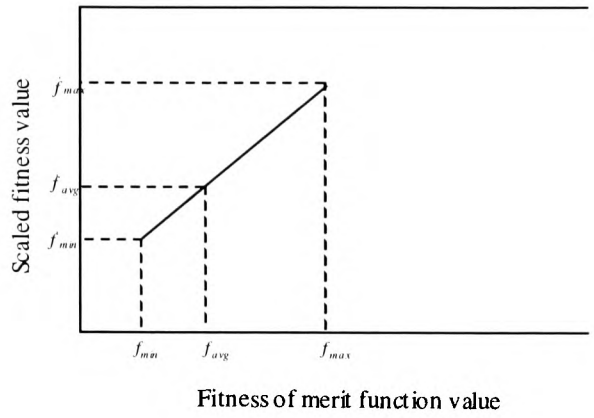


Figure 7.19 Linear fitness scaling under normal conditions.

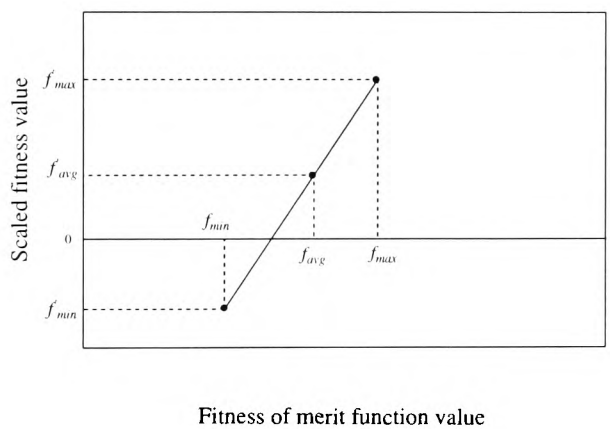


Figure 7.20 An illustration of the problem with linear scaling when f_{max} and f_{avg} are close together. This typically happens near the end of a run. Negative scaled fitness values are not allowed.

Selection scheme

- Rank selection – picks the best member each time
- Roulette wheel selection – picks an individual based on its fitness value relative to the rest of the population.
- Tournament selection – uses the roulette wheel methods to pick two candidate individuals, then pick the one with the higher fitness value. (This chooses higher valued individuals more often than roulette wheel).
- Stochastic uniform selection – picks randomly from the population.

The default options and values for parameters are shown in Table 7.18, and the results of GA optimization of PROSAIL using these parameters and the noisy dataset are shown in Table 7.19.

Parameter	*Value*
Genome representation	Real number
Algorithm type	Steady state
Scaling scheme	Linear
Selection scheme	Roulette wheel
Number of generations (n_p)	200
Population size	50
Survival probability (p_s)	0.8
Crossover probability (p_c)	0.8
Mutation probability (p_m)	0.01
Elitism	True

Table 7.18 Default parameter values for the genetic algorithm optimization

	Average retrieval error (% of parameter value)			
Parameter	*L = 0.57*	*L = 2.0*	*L = 3.6*	*Average all*
L	4.88	4.02	3.19	4.03
N	26.41	22.74	43.63	30.92
C_{ab}	10.28	3.55	11.43	8.42
C_w	5.53	8.14	11.2	8.29
C_p	31.66	23.49	27.1	27.41
C_c	30.01	27.82	31.48	29.77
$\overline{\theta}$	15.17	0.54	0.03	5.24

Table 7.19 Average retrieval error for genetic algorithm optimization of the noisy dataset using the default GA parameters given in Table 7.18, expressed as percentage of parameter value.

The results for GA with default parameters are comparable to those for SA1 using the tuned schedule. The pattern in the errors is very similar to that for SA1 with the algorithm having the most difficulty for LAI = 0.57. The errors for the most important parameters, LAI, C_{ab} and $\overline{\theta}$, are roughly equivalent to those retrieved by tuned-SA1. However, there does appear to be an improvement in the retrieval of the *difficult* parameters, N, C_p and C_c. The model is relatively insensitive to these parameters and, as a consequence, the final values of the merit functions retrieved by GA were very similar to those for tuned SA1, suggesting that both algorithms had found close approximations to the global minimum.

To determine if further improvement in retrieval accuracy could be made, the optimizations were repeated and GA parameters were varied. In general any improvements in retrieval accuracy were minor, reinforcing the view that both the default GA and tuned-SA1 algorithms had converged to a

region close to the global minimum. Figure 7.21 shows the effect on the optimization of varying the population size. As might be expected, larger populations lead to better ultimate performance, because of the larger number of schemata present in a larger population. However, it can been seen in Figure 7.21 that larger populations also have latent inertia, leading to poorer initial performance. An obvious way of increasing diversity in the population, and thereby improving the performance of the genetic algorithm, is to increase the mutation probability. Figure 7.22 shows the effect of varying the mutation probability on the test dataset and Figure 7.23 the corresponding loss of alleles. One can see from these figures that, although increasing the mutation probability results in the desired reduction in lost alleles, it also has a detrimental effect on the retrieved merit function value.

Overall the improvements in retrieval accuracy achieved by modifying the GA parameters were small: provided that the population size was larger than 200, the mutation probability was around 0.2 and the population size was at least 30. Since GAs tend to be more computationally intensive than simulated annealing, for this application they offer little advantage, provided a tuned set of schedule parameters is available for SA.

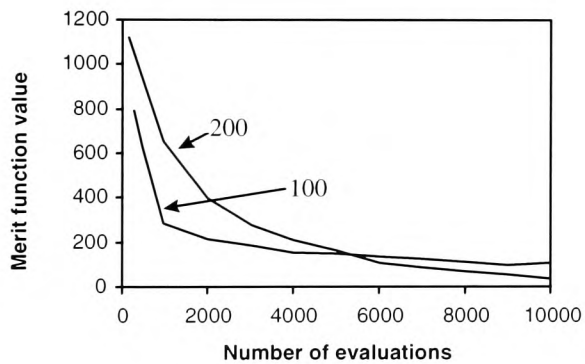


Figure 7.21 The effect of varying population size (100, 200) on the GA optimization.

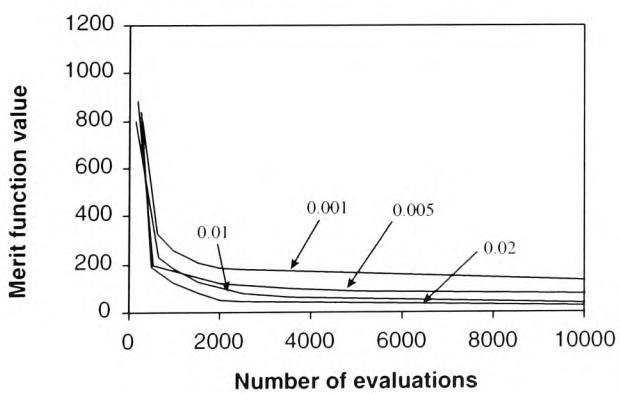


Figure 7.22 The effect of varying the mutation probability (0.001, 0.005, 0.01, 0.02) on the GA optimization

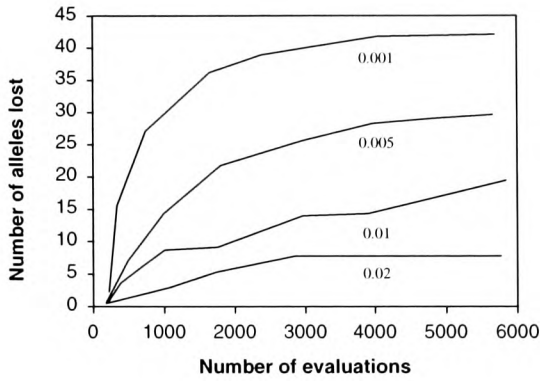


Figure 7.23 The effect of changing mutation probability (0. 001, 0.005, 0.01, 0.02) on allele loss.

7.4. INVERSION OF PROSAIL+ USING MEASURED DATA

The results using the noisy, simulated dataset illustrate that it is possible to retrieve key parameters such as LAI and C_{ab}, to acceptable levels of accuracy, under realistic conditions of sensor noise, if the PROSAIL+ model realistically represents the BRDF of the canopy. This is a "big if"; unfortunately, datasets suitable for examining the practical validity of these techniques are extremely limited.

7.4.1. Case study (Goel and Thompson 1984c)

In a series of papers (Goel and Thompson, 1984,a,b,c; Goel, Strebel and Thomspon, 1984, 1985), Goel, Thompson and Strebel investigated the possibility of determining canopy parameters by inversion of the SUITS and SAIL canopy reflectance models. The showed that the SAIL model is totally invertable (i.e. with all model parameters unknown in the inversion process) using single waveband, noise-free simulated infrared reflectance data. However, the use of noise-free simulated data does not address the problem of inadequacies in the model in representing the true canopy reflectance. Through studies with noisy simulated data, Goel and Thompson (1984b) found that they were unable to retrieve important parameters such as LAI by total inversion of SAIL. Goel and Thompson (1984c) attempted to invert the SAIL model using reflectance and bio-physical data for soybean measured by Ranson *et al.* (1985). They found that they were able to determine fairly accurately the LAI and average leaf inclination angle only by fixing the leaf and soil optical properties and the fraction of diffuse radiation (f_{sky}) at their measured values during the inversion. These results are shown in Table 7.20.

Parameter	*Measured*	*Standard deviation*	*Modelled*	*Standard deviation*	*% error*
L	2.87	0.44	3.09	0.19	7.7
$\overline{\theta}$	55.1	1.8	51.9		14

Table 7.20 Results of inversion of the SAIL model for soybean using the dataset measured by Ranson *et al.* (1985) for a fully closed soybean canopy. Leaf and soil optical properties and f_{sky} were held constant at measured values during inversion (Goel and Thompson 1984c).

From a comparison of the measured reflectance data for the soybean canopy, and canopy reflectances predicted by the SAIL model parameterized with measured parameters, one can determine that the modelled reflectances differed from the corresponding observed values by approximately 5% on average. However, detailed comparison of the measured and modelled reflectances showed that these errors were not random and varied significantly with the sun-sensor geometry, and in particular with the view zenith and azimuth angles. In general, the modelled reflectances were lower than the measured values, sometimes by up to 15%. Similar results were found by Goel and Thompson (1984c). Unsurprisingly, the best fit between the modelled and measured reflectances occurred outside the hotspot region (Chapter 3). Outside this region, and for view zenith angles of less than 25°, and view azimuth angles of between 135° and 270°, the modelled reflectances were in general agreement with the measured values. The performance of the model degraded at view zenith angles of 60° or larger. This is unlikely to prove a disadvantage for satellite derived reflectance measurements since for large view zenith angles the optical path through the atmosphere from the sensor from the target becomes very large, making atmospheric corrections difficult.

For a homogeneous canopy one might expect errors to vary with the relative azimuth angle between the sun and observer, however, dependence of view azimuth angle alone suggests that there is latent inhomogenetiy in the canopy, perhaps resulting from residual effects of the row planting scheme used for this (soybean) crop. Other errors between the modelled and measured reflectance stem from the assumptions of the model: Lambertian leaf reflectance and transmittance, the lack of a hot-spot effect, and vertical inhomgeneities in the canopy caused by the presence of stem, pods and other components.

7.4.2. Case study (Goel and Grier 1986b)

Inhomogeneous crop canopies, especially those exhibiting row-planting characteristics, are commonplace. Goel and Grier (1986a) introduced a modification to the SAIL model to allow for row effects. This row-modified model, known as ROW introduces three new parameters P - the ratio of spacing between rows and the canopy maximum height, ψ_{ROAZ} the row direction azimuth angle, and J, a parameter characterizing the canopy growth stage that is related to the percentage ground cover (G_C) (Chapter 3). Using an equivalent technique to Goel and Thompson (1984c) Goel and Grier (1986b) attempted to invert the row-modified SAIL against the Ranson *et al.* (1985) soybean dataset.

During the inversion f_{sky}, ψ_{ROAZ} and ρ_s where fixed at their measured values (Goel and Grier 1986b). Measured values for the leaf optical properties were not always available, therefore a two-step inversion procedure was used (Goel and Grier 1986b). Firstly ρ and τ were fixed at arbitrary values, or measured values where available, and the model inverted with the other parameters (LAI, μ, v, P and J) freely varying. Secondly these five parameters were fixed at the values obtained during stage 1 of the inversion and only ρ and τ allowed to vary. The results of the inversions, averaged over all datasets, are shown in Table 7.21. The percentage root-mean-square error (E_{PRMSE}) is defined by:

$$E_{PRMSE} = \left[\sum \frac{1}{N} \left(\frac{1 - R'}{R} \right)^2 \right]^{\frac{1}{2}} \tag{7.9}$$

where: R' is the measured reflectance

R is the modelled reflectance

N is the number of observations in the measured dataset

The two values of ground cover listed in Table 7.21 (G_{c1} and G_{c2}) refer to two different definitions of percentage ground cover defined by Goel and Grier (1986b), and Ranson *et al.*(1985), respectively:

$$G_{C1} = \begin{cases} 100 \dfrac{2A(JP)}{P} & J \le 0.5 \\ 100 \dfrac{2A(JP)}{P} & J \ge 0.5 \end{cases} \tag{7.10}$$

Where:

$$A(\delta) = \int_0^{\delta} f(\delta) d\delta \tag{7.11}$$

and: $f(\delta)$ is the height of the canopy at a distance δ from the centre of the canopy

$$G_{C2} = \begin{cases} 100(2J) & J \le 0.5 \\ 100 & J \ge 0.5 \end{cases} \tag{7.12}$$

The results in Table 7.21 show that the retrievals of LAI, J ,P and G_{C2} are quite good over the range of dates and that the overall E_{PRSME} is low indicating a good fit between the modelled and measured data. The value for $\bar{\theta}$ for 28 August agrees quite well with the measured values - however it should be

noted that the beta function parameters μ and ν, from which $\bar{\theta}$ was derived, were not accurately retrieved. The value of J for 28 August is not accurately retrieved, because for $J > 0.5$ the canopies overlap one another and only minor changes in the ground cover result. A similar insensitivity also explains the poor retrieval of P, the row spacing.

7.4.3. Inversion of PROSAIL+ using the "Ranson" dataset

The Ranson *et al.* (1985) soybean dataset is far from ideal for testing optimization of the PROSAIL+ model, yet it remains one of the few sets of both multi-spectral bi-directional reflectance, and bio-physical measurements for non-forest vegetation canopies. The results shown in Table 7.21 were obtained under conditions where only the LAI and LAD were unknowns during the inversion. The use of different sophisticated optimizations techniques such as simulated annealing and genetic algorithms with the PROSAIL+ model using this dataset will provide an interesting comparison of the validity and practical application of BRDF model inversion for vegetation modelling.

Once again, to simulate the measurement conditions of satellite sensor platforms, the measured datset used in the inversion was restricted to measurements lying in one plane. Because of the lack of a hot-spot representation in the PROSAIL+ model, and following on from the results with the SAIL model described earlier (Section 7.4.1), the measurement direction was chosen to lie outside the principle plane. In addition, the measured data were restricted to lie within the range 0 to 30° view zenith, and 135 to 270° view azimuth. Moreover, the original dataset was pruned to remove several observations where there were unusually large changes in the reflectance for changes in the view zenith angle. These data were affected by instrument shadow (Ranson, pers. comm.).

Date	Parameter	Modelled	Standard deviation	Measured	Standard deviation
17 Jul	L	3.6	0.23	3	0.5
	J	0.38	0.23	0.38	
	P	1.13	0.24	1.1	
	ρ	0.63	0.08		
	τ	0.3	0.1		
	$\overline{\theta}$ *	55.5	4.1		
	G_{C1}	59.2	1.1	72	
	G_{C2}	75.2	1.1	72	
	E_{PRMSE}	3.5	0.4		
24 Jul	L	4.26	0.42	3.9	0.6
	J	0.43	0.02	0.45	
	P	1.13	0.23	0.9	
	ρ	0.61	0.05		
	τ	0.34	0.06		
	$\overline{\theta}$ *	60.4	3.5		
	G_{C1}	66.7	2.3	83	
	G_{C2}	80.6	3.1	83	
	E_{PRMSE}	4.1	0.4		
28 Aug	L	3.11	0.22	2.87	0.4
	J	1.05	0.17	0.68	
	P	2.09	1.6	0.75	
	ρ	0.45		0.45	
	τ	0.52		0.32	
	$\overline{\theta}$ *	53.3	1.2	51.8	
	G_{C1}	95	0.8	99	
	G_{c2}	100	0	99	
	E_{PRMSE}	3.7	0.9		

Table 7.21 Comparison of modelled and measured row-modified SAIL model parameters from the study by Goel and Grier (1986b). The total number of observations was 238. Leaf optical properties were not measured on 17/7 and 24/7.
*The estimation of average leaf inclination angle ($\overline{\theta}$) was derived from estimations of the two beta distribution function parameters μ and ν.

The spectral directional radiometric data were measured with an Exotech Radiometer in four spectral bands, nominally corresponding to the Landsat MSS bands as shown in Table 7.22.

Exotech Band	MSS Band	Wavelength range (μ m)
1	1	0.50 - 0.60
2	3	0.60 - 0.70
3	4	0.70 - 0.80
4	5	0.80 - 1.10

Table 7.22 The four wavebands used in the Ranson *et al.* (1985) soybean dataset.

The seeding density of the soybeans was 62 kg/ha in north-south oriented rows. The row spacing was 76 cm with a mean population density of 28 plants per square meter. Agronomic and reflectance data were acquired on three dates, from July 17, July 24 and August 28, 1980. On the first two dates, the soybean canopies were incomplete with well-defined rows. By 27 August, the canopy was completely closed and exhibited minimal row structure. The sensor was mounted on a truck with an extendable boom, at a nominal altitude of 10 m above the soil surface. Measurements were acquired through all combinations of view zenith angles of 0, 7, 22, 30, 45, and 60 degree, and view azimuth angles of 0, 45, 90, 135, 180, 225, 270 and 315 degrees. Nadir soil reflectance measurements were made for a variety of sun positions (zenith and azimuth) on each of the three dates. Leaf reflectance and transmittance measurements were taken at 25 nm intervals from 450 nm to 1125 nm, in addition to a further series of measurements of the canopy components in the four Exotech bands, each for a single date. Some diffuse radiation fraction (f_{sky}) measurements were also available for Exotech bands 1 and 2 for all three dates. Some other canopy structural and bio-physical measurements were also made: these are detailed in Table 7.23.

Date	*Ave Plant Height (cm)*	*Std*	*Ave Plant Width (cm)*	*Std*	*Canopy cover G_{c2} (%)*	*Std*	*L*	*Std*	*Fresh Phytomass (g m^{-2})*	*Std*	*Dry Phytomass (g m^{-2})*	*Std*
18/7	69	4	55	5	72	4	3.0	0.5	1145	226	230	40
24/7	84	3	69	4	83	3	3.9	0.6	1540	199	320	51
28/7	102	4	104	11	99	1	2.9	0.4	2535	476	644	95

Table 7.23 Measured structural and bio-physical parameters for the soybean canopy. (Std is the standard deviation, L is leaf area index)

Results for 28 August

Table 7.24 shows the results of the SA1 optimization of the PROSAIL+ model for 28 August, averaged for six of the twelve measured datasets. Within each dataset, only measurements lying in one plane were used, with view azimuth angles between 135° and 270° and view zenith angles between 0° and 30°. Also shown in Table 7.24 are the values of the PROSPECT parameters obtained by inversion using the single-date measurements of leaf optical properties in 28 wavebands (Chapter 4). One can conclude from the results that the retrieval of LAI is quite good with a standard deviation that is well within the field measurement error. The leaf biochemical properties were not measured, so it is not possible to comment of the retrieval accuracy of the PROSPECT model parameters. However, the leaf optical properties predicted by the model using these retrieved biochemical parameters compare extremely well with the measured values of ρ and τ, with a E_{PRSME} of 2.9.

It is illustrative to compare the values for the PROSPECT parameters retrieved by the optimization algorithm using canopy reflectance measurements to those obtained by inversion using measured leaf

reflectance data (Chapter 4). As expected the values of the PROSPECT parameters retrieved by inversion of canopy reflectance data are not as good as those obtained by inversion of leaf reflectance data (Table 7.24). There are two reasons for this. Firstly the leaf reflectance data were available in 28 wavebands giving far more spectral information than the four wavebands available for canopy data. Secondly, the canopy reflectance data are composed of light that has been scattered by leaves at various levels within the canopy, and the soil background, into the field of view of the sensor. Consequently, the model is less sensitive to changes in these parameters, making their retrieval during inversion more difficult (Chapter 4).

Date	*PROSAIL Parameter*	*Inversion results (canopy data)*	*Standard deviation*	*Measured*	*Standard deviation*	*Inversion results (leaf data)*
28 Aug	L	2.99	0.29	2.87	0.40	
	N	1.73	0.24			1.39
	C_{ab}	49.98	7.3			43.5
	C_w	0.00085	0.00018			0.00064
	C_p	0.00132	0.00096			0.00080
	C_c	0.00471	0.00087			0.00153
	$\overline{\theta}$	53.6	4.4	51.8		
	E_{PRMSE}	3.6	0.41			1.01

Table 7.24 Results of SA1 optimization of the PROSAIL model using 28 August measured soybean reflectance data, averaged over sub-sets of datasets, 1,3,5,7,9,11 Ranson *et al.* (1985). In order to simulate realistic satellite-derived reflectance measurements, each subset was modified to contain reflectance measurements in one plane. Planes were chosen to have an azimuth angle in the range 135° to 270° (see text). Measured data were further restricted to lie within the view zenith range 0° to 30°, giving nine directional measurements in each of four wavebands per set. Also shown are the PROSPECT model parameters obtained by inversion of measured leaf reflectance data in 28 wavebands, 25 nm wide from 450 nm to 1125 nm

During the inversion, diffuse radiation fraction, f_{sky}, and the soil reflectance ρ_s were determined by extracting the best-fit values from a look-up table generated with the 5S and SOILSPECT+ models, respectively. The best-fit look-up table values and measured data are shown in Figure 7.24 and Figure 7.25. The fit for f_{sky} is very good indeed, remembering that the model is relatively insensitive to this parameter. The fit for soil reflectance, although slightly worse than for f_{sky}, is also very good, with errors generally of the order of 2 to 3%. This demonstrates the usefulness of the look-up table technique.

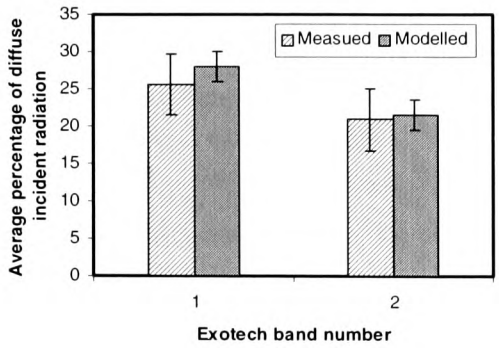


Figure 7.24 Measured average percentage of diffuse radiation for 28 August, compared with the modelled value selected from the look-up table. Note that measured data were only available for Exoctech bands 1 and 2.

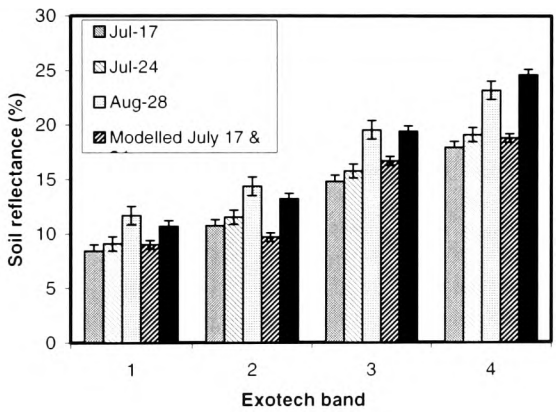


Figure 7.25 Comparison of measured soil reflectance, averaged over all solar positions for each of the three measurement dates, with modelled soil reflectance selected by the optimization algorithm from the look-up table.

Results for 17 and 24 July

On the 17 and 24 July, the soybean canopy was not fully closed (Table 7.23). In order to model such inhomogeneous canopies realistically the PROSAIL model was modified for row effects using the approach of Goel and Grier (1986a,b). As described above, this row-modified model introduces three new parameters, P, the ratio of spacing between rows and the canopy maximum height, ψ_{ROAZ} the row direction azimuth angle, and J - a parameter characterizing the canopy growth stage. The retrieval results shown below are averaged over five sets of measurements made on July 17, and six on July 24,

Date	Parameter	Modelled	Standard deviation	Measured	Standard deviation
17-Jul	L	3.4	0.22	3	0.5
	J	0.42	0.18	0.38	
	P	1.0	0.20	1.10	
	N	1.32			
	C_{ab}	45.98			
	C_w	0.00080			
	C_p	0.00124			
	C_c	0.00171			
	$\overline{\theta}$	52.9	4.6	*	
	G_{C1}	66.2	1.9	72	
	G_{C2}	74.0	1.9	72	
	E_{PRMSE}	3.54	0.4		
24-Jul	L	4.30	0.3	3.9	0.6
	J	0.48	0.02	0.45	
	P	1.10	0.23	0.9	
	N	1.90			
	C_{ab}	48.84			
	C_w	0.000105			
	C_p	0.00202			
	C_c	0.00519			
	$\overline{\theta}$	51.8	2.8	49.0	
	G_{C1}	75.7	2.7	83	
	G_{C2}	81.3	2.9	83	
	E_{PRMSE}	3.9	0.6		

Table 7.25 Results of inversion of the row-modified PROSAIL model using measured data for 17 and 24 July. Subset view angle and azimuth restrictions as in Table 7.24. ψ_{ROAZ} parameter fixed during inversion at measured value. * see text for explanation.

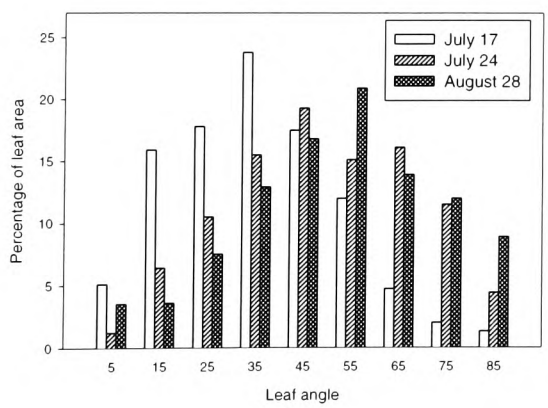


Figure 7.26 Measured variation in the soybean leaf angle distribution with date.

Table 7.25 shows the results of the inversion for the 17 and 24 July with the ψ_{ROAZ} parameter fixed at its measured value. The soil spectral reflectance and the f_{sky} parameter were determined from look-up tables as before. Once again the results are quite good, with the values of LAI, J and P all being retrieved to within accuracies comparable to the field measurement error that might be expected for these parameters. For both dates the small value of E_{PRMSE} indicates that the model fits the data relatively well. The percentage ground cover G_{c1} tends to underestimate the measure value, but G_{c2} is in quite good agreement. Although the investigators did measure the leaf angle distribution for 17 July the average leaf angle was determined as 30.2°, which seems rather too low (the measured leaf angle distribution of the three measurements dates are shown in Figure 7.26). Moreover, the investigators reported a large variation between samples, suggesting either that the measurements were in error, or that there was a large variation in leaf angle distribution throughout the canopy. Consequently, the measured average leaf angle is not listed in Table 7.25. The retrieved value of $\overline{\theta}$ for 24 July is in excellent agreement, however. Overall, the retrieved values of $\overline{\theta}$ are slightly higher than the measured values, probably because of the influence of stems on the directional reflectance. The measured leaf angle distribution did not include stems, and the PROSAIL model ignores stems, which tend to be more vertical than leaves. These factors are likely to result in a slight overestimation of average leaf angle by the model.

It is extremely unlikely that ψ_{ROAZ} will be known in practice, and therefore the optimizations were repeated with ψ_{ROAZ} free during the inversion. The results, shown in Table 7.26, indicate that LAI can still be retrieved with good accuracy, if not quite as well as when ψ_{ROAZ} is fixed. The retrieved values for the PROSPECT parameters are comparable with those in Table 7.25, however, the values for P and J are not as good. This in turn resulted in a somewhat poorer estimation of ground cover. The value for ψ_{ROAZ} itself showed a significant discrepancy with the measured value. All of these factors appear to stem from the fact that reflectance measurements in only one plane (i.e. for one effective azimuth angle) were used. The BRDF of row-canopies is characterized by a sharp decrease in reflectance as the view azimuth angle approaches the row azimuth direction (ψ_{ROAZ}), forming a *valley* in the BRDF. As canopy cover increases this decrease in reflectance lessens. When reflectance data in only one plane are available, the optimization algorithm has difficulty retrieving ψ_{ROAZ}, and the retrieval of other parameters is poorer as a consequence. When the measured dataset was altered to include additional reflectance data in other azimuth angles then retrieval accuracies improved and were comparable to those in Table 7.25. The E_{PRMSE} errors in Table 7.26 are higher than Table 7.25 and the standard deviation of E_{PRMSE} significantly higher, indicting that the inversion is unstable when ψ_{ROAZ} is free.

This may pose a problem for practical applications of this technique since it is unlikely that ψ_{ROAZ} would be known for a general vegetation target. When information about ψ_{ROAZ} is available, however, the technique looks promising. Further studies are required using more accurately measured reflectances, for a variety of different crops, before detailed conclusions may be drawn.

Date	Parameter	Modelled	Standard deviation	Measured	Standard deviation
17-Jul	L	3.7	0.54	3	0.5
	J	0.42	0.18	0.38	
	P	1.0	0.20	1.1	
	N	1.84			
	C_{ab}	49.53			
	C_w	0.00026			
	C_p	0.00319			
	C_c	0.00414			
	$\overline{\theta}$	58.5	5.1		
	ψ_{ROAZ}	165		0/180	
	G_{c1}	55.2	1.1	72	
	G_{c2}	69.7	1.1	72	
	E_{PRMSE}	5.1	0.9		
24-Jul	L	4.3	0.8	3.9	0.6
	J	0.48	0.02	0.45	
	P	1.10	0.23	0.9	
	N	2.0			
	C_{ab}	33.6			
	C_w	0.00050			
	C_p	0.00130			
	C_c	0.00303			
	$\overline{\theta}$	56.8	5.1	51.8	
	ψ_{ROAZ}	162		0	
	G_{c1}	62.0	2.3	83	
	G_{c2}	76.0	3.1	83	
	E_{PRMSE}	4.9	1.2		

Table 7.26 Results of inversion of the row-modified PROSAIL model using measured data for 17 and 24 July. Subset view angle and azimuth restrictions as in Table 7.24. ψ_{ROAZ} parameter free during inversion.

7.5. INVERSION OF THE LI-STRAHLER MODEL

The so-called Li-Strahler model (Chapter 3) is really a whole series of models developed from an original, near-nadir viewing, geometrical optical model (Li and Strahler, 1985). The basic model is designed to estimate the size and density of trees from remotely sensed images. It treats conifers as cones that cast a shadow on a contrasting background. The original model was manipulated and extended toward using the mean and variance of per-pixel reflectance to invert the model to determine

size and spacing of conifers without resolving them individually. Li and Strahler (1986) extended the nadir-viewing cone model to arbitrary illumination and view directions by adding an empirical correction for mutual shadowing and by allowing a translucent crown.

The 1985 model is invertable, but a homogeneous multi-pixel area must be identified *a priori* so that the variance in stand brightness values within the stand may be found, as described below.

The signal received by the sensor is modelled as consisting of reflected light from tree crowns, their shadows and the background with the field of view as follows:

$$S = K_g G + K_c C + K_t T + K_z Z \tag{7.13}$$

One basic problem addressed by the model is how to estimate tree crown size assuming that the areal proportions K_c and K_t can be found. Many small crowns, or a few big crowns can yield the same canopy coverage expressed by $K_c + K_t$. The separation of these effects of crown size and density is based on the relationship between image variance and tree size (Li and Strahler 1985).

To determine K_c and K_t, (and thus derive useful information about crown size and density) from the remotely sensed reflectance, the four component spectral signatures G, C, T and Z must be found. The estimation of component signatures has been one of the primary issues involved in attempting to invert the Li-Strahler model. Because of the huge variability of forest type, tree densities and illumination conditions, it is necessary to estimate the component reflectances separately for each forest type, location and series of measurements.

One method for determining spectral signatures is to use data from test stands to calibrate the model. This, of course has the great disadvantage of only being applicable where at least some field measured data are available. In theory, field measurement of spectral signatures would be the ideal way to determine G, C, T and Z. However, in practice it is extremely difficult to measure these signatures for a number of reasons. Solar illumination and other atmospheric conditions are subject to rapid changes. Therefore, it is obviously important to make measurements at near as possible to the time of the remote sensing instrument fly-over. For large areas, with several forest types and difference backgrounds all requiring detailed multiple measurements, it is clearly impossible to acquire the necessary data in a single day! Moreover, it is frequently physically extremely difficult to get above the canopy to measure the crown signatures. Such measurements usually necessitate the construction of a tower with all the associated logistical problems.

If multi-spectral measured reflectance data are used then the value of each of the component reflectances in each of the wavebands must be determined. To overcome this a linear transformation (Crist and Cicone 1984) of the original data into a two-dimensional greenness-brightness space may be employed, (Li and Strahler 1985). This linear transformation decomposes a signature into a brightness component that quantifies the overall brightness in all bands, and a greenness component that utilizes the difference between the red and near-infrared reflectance of vegetation.

Li and Strahler (1985) defined a *treeness* parameter m (provided the trees are distributed within the stand according to the Poisson distribution) as follows:

$$R^2 = \frac{V_m}{(1+W)M} \tag{7.14}$$

Where: R^2 is the mean of the squared crown radius (r^2)

M is the mean of the treeness parameter m

V_m is the variance of m

W is the coefficient of variation of the squared crown radius

The treeness parameter m is defined as the mean of nr^2, where n is the Poisson parameter. The parameter m can be thought of as a crown area index.

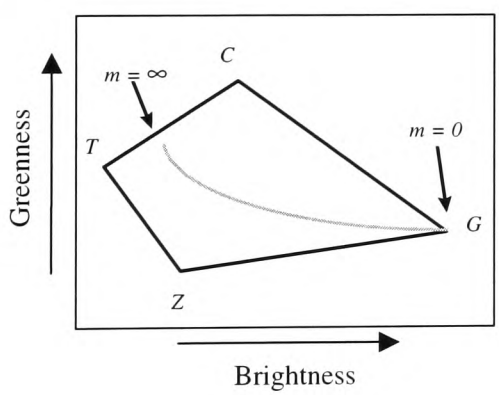


Figure 7.27 The trajectory of image values in the greenness-brightness space as the number of trees in a homogeneous stand is increased.

Figure 7.27 shows how the value of the signal, S, changes as m increases for a given stand. Illuminated background has the highest brightness value (G), and illuminated crown the highest greenness value (C). Shadowed crown in the darkest component of all, but still exhibits some greenness. When there are no trees present, $m = 0$ and the signal is just the signature of the illuminated background G. As the number of trees in the stand increases the signal becomes composed of contributions from all four components. As the number of trees becomes very large, the

background becomes completely obscured by the tree crowns; at this point, no further decrease in brightness is possible. As *m* increases, mutual shadowing of one tree by another reduces the relative proportion of shadowed background, and increases the relative proportion of shadowed crown, leading to the slight upward curvature of the line, shown in Figure 7.27.

Figure 7.27 relates *m*-values of stands to the corresponding values of the component signatures. If field measurements of *m* are available, a comparison can be made between these measured values and the *m*-values derived from the model using the greenness-brightness graph for an initial guess for the component signatures. The *best* set of component signatures can be found by minimizing the difference between the measured and modelled *m*-values. The line in Figure 7.27 represents the variation in *m* values for a homogeneous stand of trees, i.e. a stand where all tress have the same crown shape and where the background signature is uniform across the stand. For real stands, of course, there will be some degree of scatter about this line. For sparse stands the variation in background signature is dominant; while for dense stands the variation in crown shape is dominant. A two-dimensional look-up table can be constructed for the variation in *m* caused by variation in crown shape for a given set of component signatures. The look-up table is constructed by running the model forwards for every combination of values for *m* and crown shape, for a given set of values for *G*, *C*, *T* and *Z*, to provide a set of co-ordinates in the greenness-brightness plane (Figure 7.28). The table may then be used to find the *m*-value and crown shape for every pixel in the image.

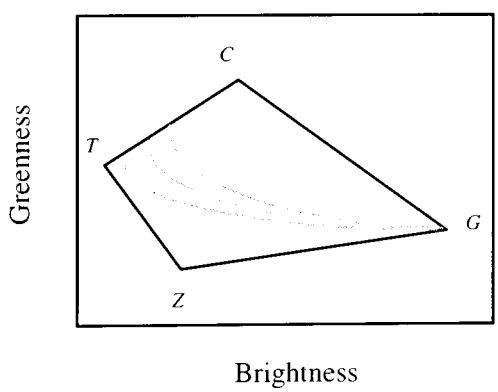


Figure 7.28 The variation of image values with *m* for different crown shapes.

The Li-Strahler model works quite well for sparse stands where trees are well separated and the shadows on crowns and background are clearly visible in the image (Li and Strahler 1988). For dense stands, little or no background is visible. The mean and variance derived from images of dense stands are contributed by the small top portion of the crowns, rather than from all four components. Because shadows play such an important rôle in the model, the results are sensitive to solar zenith angle. If the zenith angle is too small the shadows on crowns and background will be very small. Alternatively, if

the zenith angle is too large there will be too much mutual shadowing which is not catered for in the 1985 model. Li and Strahler (1992) describe an improvement to the basic model that allows for mutual shadowing explicitly, this enables the model to make more accurate predictions for large solar zenith angles and also allows non-nadir view zenith angles. However, the estimation of the four component signatures remains a serious drawback, requiring field measured data to allow inversion of the model. One possible solution to this would be to replace the translucent crowns with assemblages of leaves. Within crown scattering could then be modelled using a radiative transfer approach, parameterized by the leaf area density, leaf angle distribution and the optical properties of the leaves within the crown.

7.6. CONCLUSION

The directional information content of satellite reflectance measurements is restricted by engineering design. Satellites typically measure directional reflectance for a number of zenith angles in a single plane. Conversely, the collection of spectral information is relatively straightforward and in addition, enables the exploitation of the contrasting reflectance of vegetation and soil at different wavelengths. The PROSAIL model when combined with the SOILSPECT+ model of soil spectral reflectance, and the 5S atmospheric scattering model when modified for row effects, is capable of modelling the BRDF of soybean canopies relatively well. The parameters of PROSAIL+ are wavelength independent with the exception of the soil single scattering albedo.

The interaction amongst leaf area index, leaf optical properties and soil reflectance make the inversion of canopy reflectance models very difficult. However, the use of combined directional and spectral information, together with a look-up table for soil reflectance, appears to be a successful strategy for overcoming these difficulties.

Powerful stochastic optimization techniques such as simulated annealing and genetic algorithms are capable of inverting the PROSAIL+ model using a realistic dataset of directional and spectral reflectance measurements for a fully closed soybean canopy. The retrieval accuracy for LAI in particular was very good. SA and GA also retrieved LAI with good accuracy for an open soybean canopy by inversion of a row-modified PROSAIL model. Genetic algorithms are much less sensitive to the choice of control parameters than simulated annealing. Tuning annealing schedule parameters appears to be a process of trial and error. However, once a tuned set of parameters has been obtained the results of SA are equivalent those of the default GA. The optimum control parameters for both optimization algorithms are application specific. The availability of datasets suitable for validating this technique is extremely limited. Further, detailed, directional reflectance and physiological

measurements for crop canopies such as wheat would enable a more thorough investigation of this technique.

The Li-Strahler model, although invertable, requires field measurements of image spectral components or tree height and density for calibration purposes. This severely restricts its usefulness for modelling applications. A modification of this model to include radiative transfer within tree crowns is required.

Chapter 8

Applications of the Validation-Optimization Strategy: THE FRAMEWORK Model

Chapter 5 discussed vegetation models and the problems associated with attempting to use remotely sensed data in SVAT models at the stand scale. Two strategies for using remotely sensed data in stand-scale SVATs were described, *inversion-parameterization* and *validation-optimization* (Figure 5.4). The need for a highly flexible model that is able to integrate remotely sensed information from a number of sources and of a number of types (e.g. radar biomass estimation, BRDF data, NDVI values) was also discussed. In addition, it was noted that it is not practical to attempt to constrain models that run on a hourly time step using conventional computing techniques. Changes in canopy properties are only detectable by remote sensing instruments on a time scale of weeks. Constraining an hourly time step model over such a long period using advanced, stochastic or non-stochastic optimization techniques is normally far too computationally intensive to be practical. In this chapter, two alternative approaches to the application of the validation-optimization strategy are considered. Firstly, the application of the strategy to an empirical, daily time-step model to determine the value of the light use efficiency parameter ε, through the growing season. Secondly the coupling of empirical and mechanistic models (see Chapter 5) through the estimation of biomass allocation.

In general, SVAT models can be classified into big-leaf or multi-layer models depending on their assumptions about canopy structure. The representation of vegetation function can range from a highly empirical to a detailed mechanistic description and the time-step of the model can vary from a fraction of an hour to several days, depending on the application and how the processes of photosynthesis and growth are modelled. This chapter describes the development of a coupled SVAT-BRDF model that is flexible enough to be parameterized for almost any plant type in any location. In addition it may be configured as an empirical model, a mechanistic model, or both, use big-leaf or multi-layer representations, and be parameterized or constrained using remotely sensed data from almost any source through built-in stochastic optimization. Through the use of modular programming techniques, the representations of vegetation function can be readily interchanged without affecting other components of the model. For example, the photosynthesis routine may be easily swapped from a mechanistic Farquhar-type C_3 model to an empirical C_4 model with minimal effects on the rest of the SVAT. In summary, the main features of the model are:

- big-leaf or multi-layer canopy representation,
- any number of canopy layers,
- any time step – time step may be continuously varied,
- complete freedom for specifying complexity of functional representation,
- modular design enables representation to be easily altered and enables configuration of the model for any vegetation type and geographical location,
- coupled BRDF model for prediction of directional reflectances and calculation of APAR, and
- an integrated suite of optimization routines for BRDF model inversion, or optimization of parameter trajectories.

Since it is usual practice to give new SVAT models a name (if only to make referring to them somewhat easier), the epithet FRAMEWORK has been chosen, to reflect the flexibility of the model.

8.1. OBJECT-ORIENTED PROGRAMMING

The key to developing a model with the features described above is the use of object-oriented software design. Object orientation is the use of objects and classes in analysis, design, and programming. The use of objects distinguishes object orientation from traditional structured methods. Objects provide a canonical focus throughout analysis, design, and implementation by emphasizing the state, behaviour, and interaction of objects in models. This building block approach provides a significant improvement in basic component definition and interfacing.

The main practical benefits of adopting an object-oriented approach are:

- faster development ,
- easier maintenance ,
- more easily modifiable code ,
- reuse of software and designs, and
- systems more change resilient and more readily evolved.

The ability to reuse software and designs is an important benefit of object orientation. The definition of an object can be easily altered with minimum consequent effects on other objects. With incremental

modification, one can make extensions or modifications to a system's behaviour without modifying existing code but rather by adding new code. Resistance to change is a valuable property: system objects change infrequently while processes and procedures are frequently changed, providing object-oriented systems with more resilient system organization.

Objects

An object represents an individual, identifiable entity, either real or abstract, with a well-defined rôle in the problem domain. An object has a state, behaviour, and identity associated with it. Objects with a similar structure and behaviour form a common class. Most object-oriented languages provide a well-defined interface to their objects through classes. Within this context, a *class* is a specification of structure (instance variables), behaviour (methods) and inheritance (parent - child relationships) for objects. An object class is a set of objects that share a common structure and a common behaviour (Booch 1991). By allowing the programmer to define object classes, he or she can define his or her own kinds of objects, and instantiate them (i.e. create an instance) as needed. An instance of a class is an object that is described by that class. Instantiation of a class is the creation of a new instance of that class. An object class specifies a set of methods visible to the programmer, a set of hidden instance variables and a set of hidden operations that implement the methods. The instance variables can only be modified indirectly by invoking the operations. When a new instance of an object class is created, it has its own set of instance variables, and it shares the method implementations with other instances of its class.

Methods

Methods implement behaviour, i.e. how an object acts and reacts, in terms of its state changes and message passing. Each class has an associated group of methods, in the form of functions or procedures. Typically, methods access the internal state of an object of that class to perform some sort of operation.

Three main principles govern object-oriented software design:

- encapsulation or data hiding,
- inheritance, and
- polymorphism.

Encapsulation

Encapsulation is the process of hiding all details of an object that do not contribute to its essential characteristics. Consequently, encapsulation hides the implementation details of the object and the only thing that remains externally visible is the interface of the object, i.e. the set of all messages the object can respond to. Once an object is encapsulated, its implementation details are not immediately accessible, instead they are only indirectly accessible *via* the interface of the object

Inheritance

Inheritance is an inter-class relationship where one class forms the parent and another the child class. The child inherits the basic characteristics of the parent. A child may modify the characteristics inherited from the parent and introduce new characteristics of its own. Inheritance enables programming by extension as opposed to programming by reinvention. Inheritance allows the commonality of objects to be used to advantage in modelling and constructing object-oriented systems. Inheritance also provides for code and structural reuse. Class libraries also allow reuse between applications, potentially allowing order-of-magnitude increases in productivity and reductions in defect rates (program errors), as library classes have already been tested and further use provides further testing, providing even greater reliability.

A consequence of inheritance is that a class does not have to be modified if it is close to what is required; a derived class can be created to specialize it. This avoids code redundancy, since otherwise code would have to be copied and modified. Thus, inheritance is an incremental modification mechanism, i.e. a code-reuse mechanism to build new objects incrementally out of old ones. With inheritance, code and structure can be shared between different objects.

Polymorphism

Polymorphic functions can be of many different types, i.e. serve the same purpose and perform the same actions on variables of different types, but the actual coding required to perform the these actions may differ between types. Two types of polymorphism can be identified, namely parametric and inclusion polymorphism. Parametric polymorphism is embodied by generic functions that need a parameter to know with which type they should actually be used. Inclusion polymorphism occurs through inheritance, when an object may be viewed as belonging to many different classes.

The reader is referred to Sequeira, Olson and McKinion (1997), Lemmon and Chuck (1997), Acock and Reddy (1997) and Reynolds and Acock (1997) for an excellent introduction to introduction to the use of object oriented programming techniques in vegetation modelling. These papers comprise a special edition of Ecological Modelling 94 (1997) "Modularity in Plants".

8.2. THE FRAMEWORK MODEL

By adopting object-oriented software development techniques and through careful design it is possible to build an extremely high degree of flexibility into SVAT models. This is exemplified by the FRAMEWORK model developed for this study. In conventional SVATs, attempts are made to represent the functioning of individual vegetation types, e.g. forests in either a relatively empirical or a

more mechanistic fashion. The assumptions, and approximations made in the model, are entrained within the program code of model itself, and although, for example, it is normally possible to parameterize forest models for different tree species in temperate zones, the same model may not work well if parameterized for C_4 crops in semi-arid regions.

Unfortunately, the full benefits of object derivation and inheritance cannot be used within FRAMEWORK because of the need to keep the model as general as possible. In order to use derived objects the interface for each child object must be fixed. This is so that all the parameters required by the methods are available and within *scope* (i.e. are available to be accessed by the program). The number of ways of implementing the methods for objects such as photosynthesis is very large. It would be too restrictive to define the parameters allowed for photosynthesis methods within the object hierarchy itself, because there may be little commonality between two different representations of the same method. The Farquhar model, for example, requires a very different set of parameters from a simple empirical model. Strict adherence to object design philosophy would necessitate that all conceivable external parameters that may be required by any photosynthesis model are included in the interface for this object. This is not only cumbersome, but it is also not possible to predict the parameters that may be required by a photosynthesis model that has not yet been devised.

If interface specifications do not exist the reusability of code is compromised because the programmer may have to re-code existing objects to allow appropriate variable to be passed to newly introduced objects. The interfaces within FRAMEWORK offer a compromise solution to this problem, which is to define general interfaces for all objects, but to provide ancillary code to allow new objects, methods and interfaces to be added in a manner that complements existing ones. These general interfaces are intended to provide a base upon which to build new objects through the use of inheritance. New member variables, not provided by the default objects may be introduced by deriving a new child object from a default one. Consider the following hypothetical example: a programmer wants to introduce a new photosynthesis object that contains new methods and requires a additional variable *a leaf cuticle wax thickness index*, not provided in the general interface specification. The structure object would be the logical object to contain this variable, therefore a new structure object would be derived inheriting the existing variables and methods from the original, and providing new methods for inputting, calculating and returning the *leaf cuticle wax thickness index*, w_{ax}. In order to w_{ax} to the new photosynthesis object, the calculation object would have to be modified. This latter step would, of course, not be required if all interface specifications were fixed. However, in terms of vegetation modelling the added flexibility of the approach used in FRAMEWORK outweighs the benefits of fixed interfaces.

The design philosophy used for FRAMEWORK was intended to not only provide software to model vegetation function, but also to provide a structure where self contained modules, each representing a specific vegetation function, can inter-communicate. Thus the model becomes an assemblage of objects and associated methods, represented by the modules, with communication between objects handled by a system of software interfaces – hence the name FRAMEWORK. In such a system it is extremely easy to change radically the representation of specific vegetation functions by changing the appropriate objects. This may be done in order to reconfigure the model for different vegetation types, e.g. C_4 instead of C_3 photosynthesis, or may be used to examine the effect of simplifications in the representation of individual plant functions on the overall model behaviour. The highly modular structure of FRAMEWORK extends to the representation of canopy structure so that the model may be configured as either a big-leaf or a multi-layer SVAT. The choice of time step is also completely configurable and may be varied during the model run. In addition to modelling vegetation function, the model structure facilitates the integration of remotely sensed data to the model through either the inversion-parameterization or the validation-optimization approaches described in Chapter 5.

FRAMEWORK is not a rapid application development tool and contains no user interface. FRAMEWORK is designed for maximum flexibility and performance and all input and output is handled via text files. The user requires a working knowledge of objected-oriented programming techniques and the C++ language in order to configure the model.

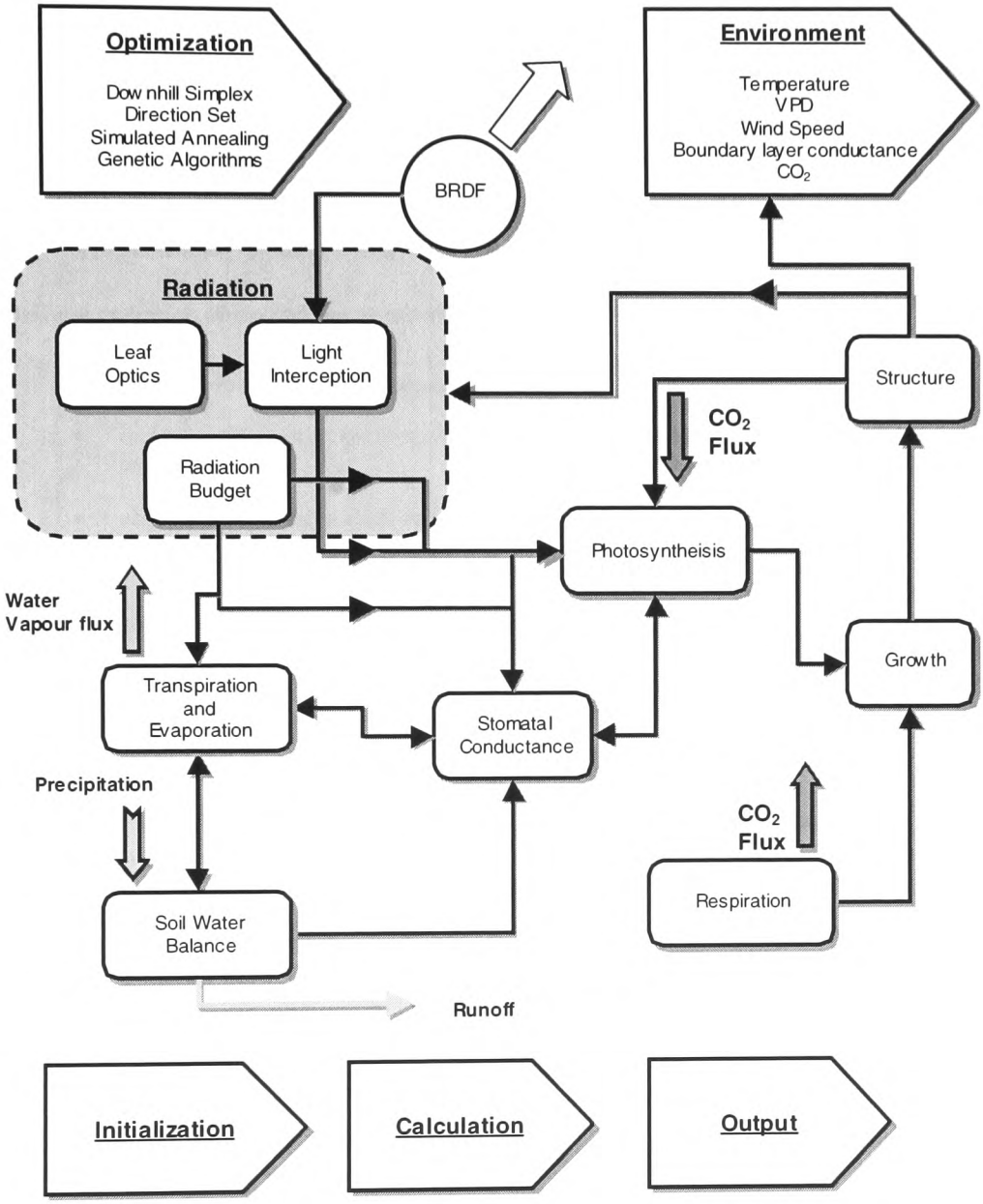


Figure 8.1 A schematic representation of the interactivity of the default FRAMWORK objects. Rectangular boxes indicate objects and their associated methods. The radiation object is a parent with three child objects: leaf optics, light interception and radiation budget. Arrows linking objects represent the transfer of data through object interfaces. The tag-shaped boxes indicate that these objects (optimization, environment, initialization, calculation and output) interact with all the other objects in the model.

With such a flexible structure, a wide range of configurations of FRAMEWORK is possible. Figure 8.1 shows one typical configuration that is used as the default: the boxes represent objects and associated methods, and the arrows the flow of data through object interfaces. In addition, an object to read in configuration information, and an object that specifies site specific variables such at latitude, longitude and elevation above sea-level exist but are not shown in Figure 8.1. An object stores data (information) in the form of member-variables and the methods are called to perform operations (calculations) on these stored data. This concept is best explained through an example.

Consider a **radiation** object, whose member-variables are:

- total incident short wave radiation
- incident radiation in photosynthetically active radiation (PAR) and near infrared (NIR) wavelengths
- absorbed radiation in PAR and NIR
- total absorbed short wave + long wave radiation
- ratio of diffuse to total incident radiation in PAR and NIR

The object methods are as follows:

- *Update* – reads the meteorological input data file and updates the total incident short wave radiation; calculates incident PAR and NIR, and updates these variables.
- *Calculate absorbed radiation* – calculates the absorbed radiation in a given waveband. Note that if a **BRDF** object exists and is to be used for calculating APAR, this method provides an interface between the **radiation** object and the **BRDF** object.
- *Calculate net radiation* – uses standard equations.

The model operation proceeds in three stages:

Initialization

The **initialization** object is called and reads in its own configuration file. This file indicates which objects are to be constructed for this particular overall model configuration, i.e. it lists the objects required by the model. These objects are then individually initialized *via* calls to their constructors. Normally the constructors read in appropriate configuration information from specified files, but the designer has complete freedom over the design of the construction process.

Processing

Following initialization, the **calculation** object is called to perform the model iterations based on the specified time-step. For specific modelling objectives it is likely that this object will have to be

customized by the user, but several basic model configurations and their associated **calculation** objects are included in the basic model. For LUE models the calculation object may simply consist of calling the appropriate **radiation** method to calculate APAR followed by the **growth** and **structure** methods to calculate any corresponding change in canopy architecture. More complex mechanistic-type models may involve a much more complex processing stage. A typical processing stage for a mechanistic model is shown in Figure 8.2.

Data output

With so many different configurations of the model possible, the number of combinations of output parameters that might be requested by the user is almost limitless. Since flexibility was the raison d'être of the model great care was taken in the design of output methods. This significantly increased the amount of coding required to develop the model but was a vital part of the modelling approach adopted. The model stores all variables internally in non-derived units (e.g., assimilation is stored as mol m^{-2} s^{-1}). Separate methods are supplied to return the values of internal data members of every object, this enforces the object-oriented paradigms of data encapsulation and hiding, yet allows any of the model variables to be output on request. The **output** object is called from the **calculation** object to request data output, for any time-step, to screen or file. Unfortunately, for specific applications, to maintain maximum flexibility the **output** object has to be hand-coded by the user. This is a relatively simple task: templates are supplied for commonly used output formats, and standard unit conversions are incorporated in the object methods. All output is in the form of delimited text files: no graphical output is possible. This is not a major limitation, since the data are easily loaded into any graphical analysis package for processing and display.

Incorporating Remotely Sensed Data

The model is designed so that remotely sensed data may be incorporated at any time, through the use of the inversion-parameterization or validation-optimization techniques, described in Chapter 5. A **schedule** object contains information on the date, time and type of the remotely sensed data available, and the optimization strategy to be used. The **schedule** object is not dynamically configurable, and must be customized by the user in co-ordination with the **calculation** object before a model run. Again, to ease customization, several configurations are hard-coded as shown in Table 8.1.

Big-leaf v Multi-layer Representation

Any number of layers or compartments can be defined in the model. This is achieved through the simple expedient of defining multiple **structure** objects. A big-leaf model requires only one **structure** object, whereas an n-layer model (or one that has n components) would normally be configured with $n + 1$ **structure** objects, as shown in Figure 8.3.

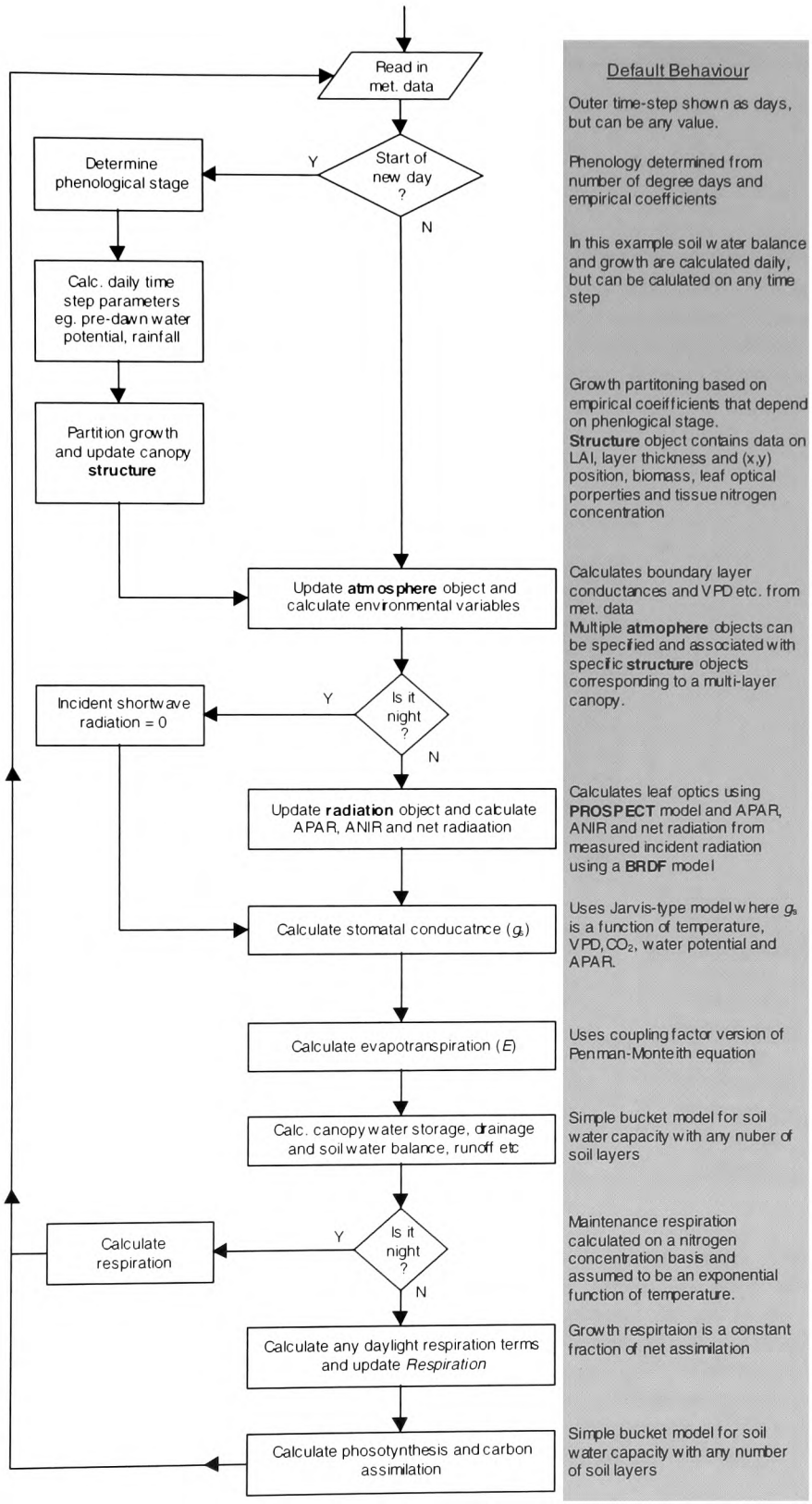


Figure 8.2 The processing stage of the **calculation** object

Remotely Sensed Data	*Inversion - Parameterization*	*Validation - Optimization*
Single–date directional reflectance	✓	
Directional reflectance series		✓
NDVI	✓	✓
LAI	✓	✓
Biomass	✓	✓
CO_2 flux		✓

Table 8.1 Default methods for incorporating remotely sensed data into the FRAMEWORK model. A tick in the inversion-parameterization column indicates that the corresponding data may be used to invert FRAMEWORK to derive model parameters. A tick in the validation-optimization column indicates that the data may be used to constrain the *state* of the model over time.

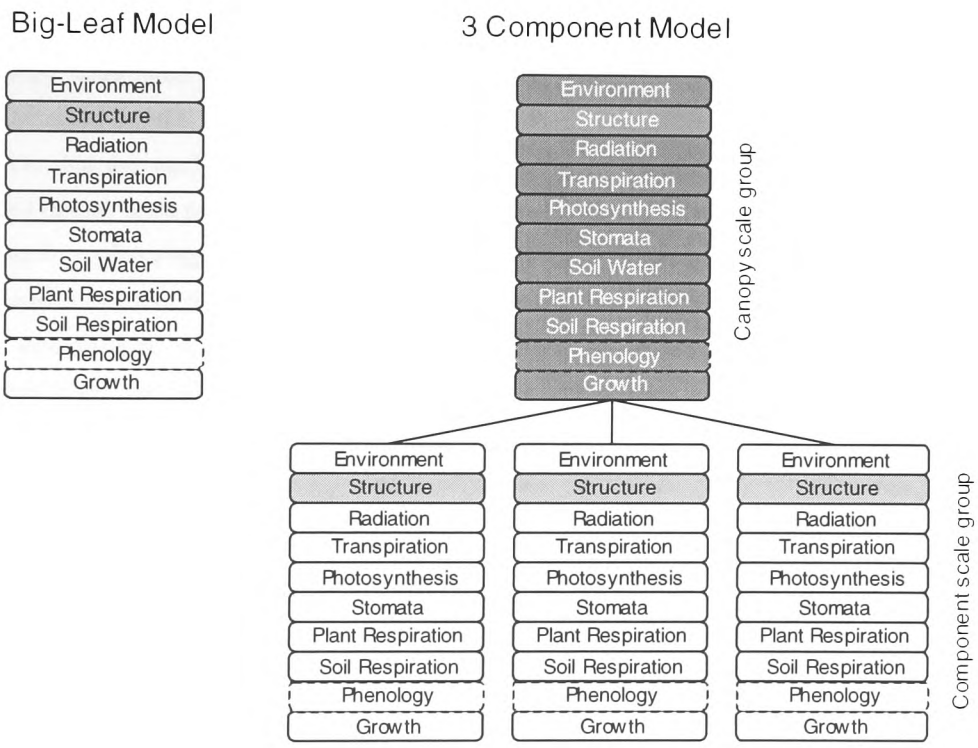


Figure 8.3 An object model diagram of a big-leaf model and a multi-component model. The multiple components may represent layers at different height within a canopy, or canopy sub-volumes with different, horizontal and vertical positions within the canopy. Alternatively, components may represent different individual plant within a landscape (e.g. individual trees in a forest stand). The big-leaf model consists of one group of objects. The three component model consists of four groups, one group for each of the canopy components (the component-scale objects) and one for the canopy itself (the canopy-scale objects). Although the canopy scale group is composed of the same objects, different methods may be employed.

The *extra* **structure** object in the multi-component model is used to store canopy-scale variables such as the canopy carbon assimilation calculated from the weighted sum of the carbon assimilation from each model component. In addition to the canopy structure, the environmental conditions, especially the radiation environment, will be different for each component (otherwise, a big-leaf model would suffice!). Consequently, each **structure** object is associated with an **environment** and a **radiation** object, together with appropriate objects to calculate and store evapotransipration, stomatal conductance, respiration and photosynthesis for each component. Growth may be modelled at the canopy scale only or, as in Figure 8.3, for each component separately. A single **soil** object is usually associated with the canopy-scale object group, but may alternatively be associated with the lowest canopy layer, for example.

The existence of a canopy scale group is a convenient way of summing component level parameters into canopy scale ones: however, it also facilitates the linkage between big-leaf and mechanistic models required by the validation optimization approach. Consequently, it is the canopy object parameters that are used in any optimizations, and user-specified mappings are used to distribute optimized canopy scale parameters amongst the component objects. This mapping can be dynamic if, for example, a **phenology** object is associated with each model component. Although the canopy scale group consists of the same objects as the component scale groups the object variables will, of course, have different values, and different object methods may be employed that are appropriate to the canopy scale. Interactions between component-level objects may occur, for example the incident radiation for one layer in a canopy will typically depend on the radiation absorbed and scattered by other layers higher in the canopy. The way in which such interactions are handled depends upon the specific implementation of the individual processes. Taking the radiation example, typically each component will calculate its radiation regime by obtaining transmitted radiation data from its neighbouring components using the component-scale and canopy-scale structure objects variables to determine relative positions.

Optimization Suite

The optimization suite forms an integral part of the model, although it can also be used as a stand-alone application. Four optimization techniques are included in the suite, Downhill Simplex, Direction Set, Simulated Annealing (2 versions) and Genetic Algorithms; these are described in detail in Chapters 6 and 7. The optimization techniques effectively attempt to optimize a user-supplied function with an array of variables and an array of parameters. The parameters represent model variables which, although not fixed, do not vary with respect to the optimization. A simple example is the solar zenith angle which may vary between observations but is not varied within the optimization.

In addition some model variables may not be considered as free during optimization. An example of this might be leaf reflectance that has been fixed in some way, probably through field measurement.

8.3. DEFAULT METHODS

Although FRAMEWORK is designed to be completely configurable, both in the representation of vegetation functioning, time step and through the use of remotely sensed data as a constraint, a number of default objects and methods are included in the model. The following section discusses these defaults methods.

8.3.1. Optimization

The optimization object contains methods for four optimization techniques, Downhill Simplex, Direction Set, Simulated Annealing and Genetic Algorithms. These are discussed in detail in Chapter 6, and therefore will not be considered here.

8.3.2. Stomatal Conductance

Stomatal conductance for water vapour g_s (mol m^{-2} s^{-1}) is modelled using a Jarvis-type approach (Jarvis 1976):

$$g_s = g_s(Q_a) \cdot g_s(T) \cdot g_s(D_v) \cdot g_s(\Psi) \tag{8.1}$$

Where Q_a is absorbed photosynthetic photon flux density (PPFD)

T is leaf temperature

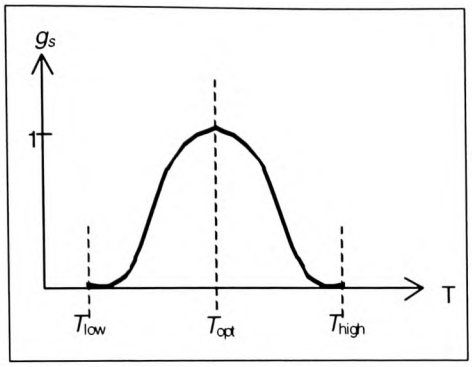
D_v is saturation vapour pressure deficit

Ψ is soil water potential

Temperature dependence

g_s is assumed to be a function of temperature given by (Jarvis 1976):

$$g_s(T) = \frac{(T - T_{low})(T_{high} - T_{opt})^{\frac{T_{high} - T_{opt}}{T_{opt} - T_{low}}}}{(T_{opt} - T_{low})(T_{high} - T_{opt})^{\frac{T_{high} - T_{opt}}{T_{opt} - T_{low}}}} \tag{8.2}$$



Where: T is the leaf temperatue

T_{opt} is the value of T at which g_s is a maximum (the optimum temperature)

T_{low} is the low temperature threshold

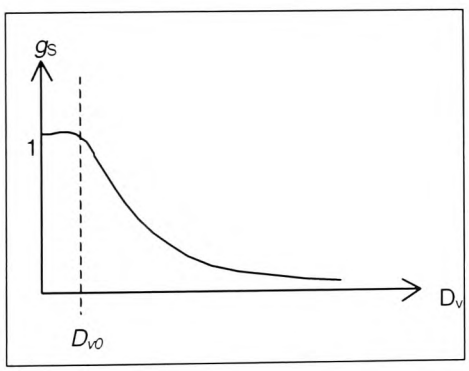
T_{high} is the high temperature threshold

Figure 8.4 The dependence of stomatal conductance on temperature.

VPD dependence

g_s is assumed to exhibit a reciprocal relationship with increasing VPD above a given threshold:

$$g_s(D_v) = \frac{1}{\left(1 + k_{g_s(D_v)}\right) \cdot \left(D_v - D_{v0}\right)} \tag{8.3}$$



Where: $k_{g_s(D_v)}$ is a reciprocal coefficient

D_{v0} is the VPD threshold

Figure 8.5 The dependence of stomatal conductance on vapour pressure deficit.

Leaf water potential dependence

The dependence of g_s on soil water is modelled as a linear function of the water potential (Jarvis 1976), which is in turn modelled as a function of the available soil water (Figure 8.6 and Figure 8.7). The following parameters must be specified: $\Psi_{intercept}$ (Pa), Ψ_{min} (Pa), $i_{threshold}$ (Pa), $g_{smin}(\Psi)$ (dimensionless).

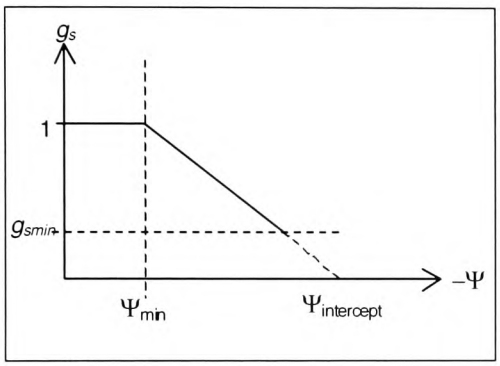


Figure 8.6 The relationship between stomatal conductance and water potential.

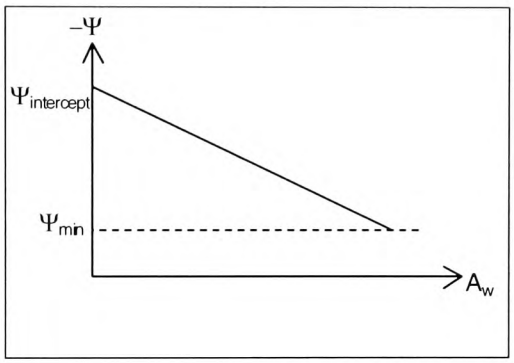


Figure 8.7 The relationship between water potential and percentage of available soil water (A_w)

PAR dependence

g_s is related to absorbed PAR Q_a by a rectangular hyperbola (Jarvis 1976):

$$g_s(Q_a) = \frac{g_{s\max}\delta_{g_s}\left(Q_a + \left(\frac{g_{sdark}}{\delta_{g_s}}\right)\right)}{g_{s\max} + \delta_{g_s}\left(Q_a + \left(\frac{g_{sdark}}{\delta_{g_s}}\right)\right)} \tag{8.4}$$

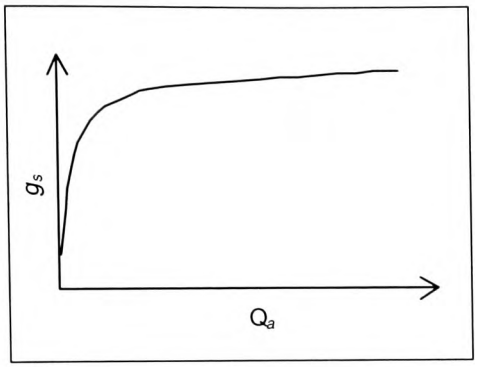


Figure 8.8 The response of stomatal conductance to APAR.

Where: g_{smax} is the maximum allowed value for g_s

g_{sdark} is the stomatal conductance in the dark

δ_{g_s} is the initial slope of the g_s v light curve.

8.3.3. Soil Water

The soil is modelled as a collection of n layers each with a specified water holding capacity. The top layer receives water input in the form of free through fall precipitation and drainage of water from leaves. Precipitation is partitioned into free throughfall, interception and retention by leaves. A proportion of the water intercepted by the leaves leaves runs off to form stem flow. Water held on the canopy evaporates and runs off if the water holding capacity of the leaves is exceeded. When the water holding capacity of each soil layer is exceeded the water drains into the next layer down. Water is lost from the layer upper (and optionally lower layers) through transpiration demand. Water may also evaporate from the soil surface.

Free throughfall is calculated from :

$$P_{fff} = P_g p_{fff} e^{-L\cos(\bar{\theta})} \tag{8.5}$$

So that the input of water onto the canopy (P_{can}) is given by:

$$P_{can} = P_g - P_{fff} - P_{stem} \tag{8.6}$$

Where: P_{ftf} is the amount of precipitation that forms free throughfall

P_g is gross precipitation (mol m^{-2} ground area)

P_{can} is the increase in water stored on the canopy (mol m^{-2} ground area)

P_{stem} is the proportion of the water intercepted by the canopy that forms stem flow

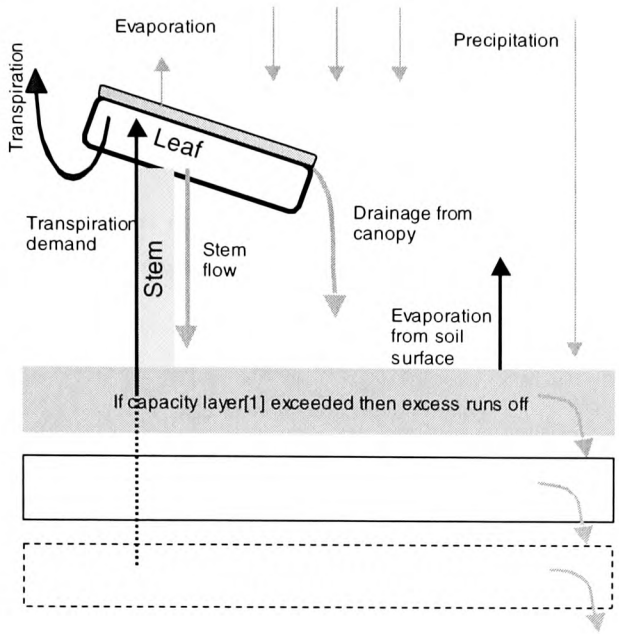


Figure 8.9 The water balance submodel.

The available water (Figure 8.7 and Figure 8.8) is the sum of the water content of all n layers from which the plant can extract water (Figure 8.9). The soil water potential, Ψ (in Pa) is calculated from:

$$\Psi = \Psi_{\text{int ercept}} + A_w \left(\frac{\Psi_{\min} - \Psi_{\text{int ercept}}}{W_{cap}} \right) \tag{8.7}$$

Where: A_w is the available soil water per unit area

W_{cap} is the water capacity of the layers from which the plant can extract water.

$\Psi_{\text{int ercept}}$ and $\Psi_{\min}$ are defined in Figure 8.6 and Figure 8.7.

8.3.4. Evapotranspiration

The calculation of evaporation of surface water stored on leaves and transpiration of water vapour through the leaf stomata are considered separately as follows.

Transpiration is modelled using the Jarvis and McNaughton (1986) "Omega" formulation of the Penman-Monteith equation. The exact form of the equations varies slightly for multi-component

amphistomatous and hypostomatous model configurations, and for big-leaf models; these variations are detailed in Appendix C. The **transpiration** object contains methods for all the configurations, which may be selected by means of a software switch.

Evaporation from the wet surfaces of leaves is calculated using the Penman equation, modified for isothermal net radiation, R_{iso} (Monteith and Unsworth 1990):

$$\lambda E = \frac{\varepsilon R_{iso} + \frac{\rho c_p}{\gamma} D_v g_R}{\varepsilon + 1 + \frac{g_R}{g_b}} \tag{8.8}$$

Where: ρ is density of dry air

c_p is the molar heat capacity of dry air at constant pressure

λ is the molar latent heat of vapourization of water

γ is the psychrometric constant ($c_p P/\lambda$)

D_v is the water vapour saturation deficit of air

ε is the increase in latent heat content per increase in sensible heat content of saturated air at ambient temperature (s/γ), s is the slope of the curve relating saturation vapour pressure to temperature

g_R is the radiative conductance

g_b is leaf boundary layer conductance

Increasing water stored on the canopy surfaces increases evaporation and decreases transpiration in a linear manner. The potential maximum rates of transpiration (E_{trans}^{pot}) and evaporation (E_{evap}^{pot}) are calculated as described above, the actual rates being weighted by the amount of water stored on the canopy (P_{can}) relative to the canopy storage capacity ($P_{can}^{\max}$):

$$E_{evap} = E_{evap}^{pot} \frac{P_{can}}{P_{can}^{\max}}$$

$$E_{trans} = E_{trans}^{pot} \left(1 - \frac{P_{can}}{P_{can}^{\max}} \right) \tag{8.9}$$

8.3.5. C_3 Photosynthesis

The Farquhar and von Caemmerer (1982) mechanistic photosynthesis model (FvC) is used by default. A detailed description of this model is given in Appendix B. The dependence of photosynthesis on ambient CO_2 concentration is inherent in the FvC model. The sensitivity of assimilation to temperature arises from the dependence of the parameters the Michaelis-Menton coefficient for photosynthesis, K_m, the maximum carboxylation rate of rubisco, V_{cmax}, and the maximum electron transport rate, J_{max}, on temperature. The effect of nitrogen on photosynthesis is introduced through a linear dependency of J_{max} and V_{cmax} on leaf nitrogen content. A

8.3.6. C_4 Photosynthesis

Several models of C_4 photosynthesis have been proposed, all of which link a C_3 photosynthesis model with a carbon pump driven by the activity of PEP carboxylase (e.g. Berry and Farquhar 1978). Carbon, derived from intercellular CO_2 is fixed into C_4 acids in the mesophyll, transported to the bundle sheath cells, and released as CO_2. Leakage of inorganic carbon from the bundle sheath cells back into the intercellular spaces occurs because of the large gradient in CO_2 concentration created by the metabolic pump. The steady-state balance between CO_2 concentration and leakage can be expressed as (Berry and Farquhar 1978):

$$A_n = (W_p - L) - R_d \tag{8.10}$$

Where: A_n is the net rate of C_3 photosynthesis in the bundle sheath cells

W_p is the velocity of the PEP carboxylase metabolic pump

L is the leakage flux of CO_2 from the bundle sheath cells

R_d is the temperature dependent rate of leaf respiration

Equation 8.10 forms the basis of the intercellular transport model (ICT) described in Appendix A of Collatz *et al* (1992). The following sections use the Collatz *et al* (1992) notation.

The assimilation rate A is assumed to be a function of the CO_2 and O_2 partial pressures in the bundle sheath cells. The formulation of Farquhar, von Caemmerer and Berry (1980) can be used to express the C_3 photosynthesis in the bundle sheaths as rubisco-limited and a light limited terms:

Rubisco limited

$$A = \frac{V_{cmax}\left(p_{bs} - \dfrac{0.5O_{bs}}{\tau}\right)}{p_{bs} + K_c\left(1 + \dfrac{O_{bs}}{K_o}\right)} \tag{8.11}$$

Light-limited

$$A = a\alpha_r f Q_p \frac{p_{bs} - \frac{0.5O_{bs}}{\tau}}{p_{bs} + \frac{1.17O_{bs}}{\tau}} \tag{8.12}$$

Where the coefficient 1.17 reflects the extra light-generated ATP required by phototrespiration (Farquhar and Berry 1978),

Where: V_{cmax} is the maximum capacity for CO_2 fixation by rubisco

K_c, K_o are the Michaelis-Menton coefficients for carboxylation and oxygenation by rubisco respectively

p_{bs}, O_{bs} are the bundle sheath partial pressures of CO_2 and O_2 respectively

ι is the specificity of rubisco for CO_2 relative to O_2

a is the leaf quantum absorptance

α_r is the intrinsic quantum yield of C_3 photosynthesis

f is the fraction of absorbed photons used for C_3 photosynthesis

Q_p is the incident quantum flux density

The CO_2 and light dependence of the activity of PEP carboxylase (W_p) are given by

CO_2 limited

$$W_p = k_p \frac{p_i}{P} \tag{8.13}$$

Light limited

$$W_p = a\alpha_p (1 - f) Q_p \tag{8.14}$$

Where: p_i is the intercellular CO_2 concentration

k_p is the apparent first-order rate constant for PEP carboxylase with respect to p_i (theoretically equal to V_{cmax}/K_m) and is derived from the initial slope of the leaf CO_2 response curve.

α_p is the reciprocal of the quantum requirement for the production of PEP

The leakage of CO_2 from the bundle sheath to the mesophyll is given by:

$$L = \frac{(p_{bs} - p_i)}{r_c P} \tag{8.15}$$

Where: r_c is the resistance to CO_2 diffusion from the bundle sheath cells

The rate of O_2 production and its diffusion from the bundle sheath into the mesophyll is given by:

$$\lambda A = \frac{(O_{bs} - O_i)}{PDr_c} \tag{8.16}$$

Where λ is the proportion of total net O_2 production occurring in the bundle sheath

O_i is the partial pressure of O_2 in the intercellular spaces

D is a constant describing the differing diffusivities of CO_2 and O_2.

Solving equations 8.15 and 8.16 for p_{bs} and O_{bs} respectively, and substituting into equations 8.11 and 8.13, or 8.12 and 8.14 and finially combining with equation 8.10 yeilds a quadratic equation for either the CO_2 limiting or light limiting value of A_n as a function of p_i, O_i and Q_p of the form:

$$aA^2 + bA + c = 0 \tag{8.17}$$

Where the coefficients a,b and c are given by:

CO_2-limiting

$$a = P\left(\frac{\lambda r_c DK_c}{K_o} - r_c\right)$$

$$b = \mathrm{Pr}_c\, p_i + K_c + \frac{O_i K_c}{K_o} + \frac{0.5\lambda r_c PDV_{cmax}}{\tau} + V_{cmax} r_c P \tag{8.18}$$

$$c = V_{cmax}\left(\frac{0.5O_i}{\tau} - r_c Pk_p p_i - p_i\right)$$

Light-limiting

$$a = P\left(\frac{1.17\lambda DK_c}{K_o} - r_c\right)$$

$$b = a\alpha_p(1-f)r_cPQ_p + p_i + \frac{1.17O_i}{\tau} + \frac{0.5a\alpha_r fQ_p\lambda r_cPD}{\tau} + a\alpha_r fQ_p r_cP \quad (8.19)$$

$$c = a\alpha_r fQ_p\left(\frac{0.5O_i}{\tau} - a\alpha_p(1-f)r_cPQ_p - p_i\right)$$

Temperature sensitivity

The temperature sensitivies of the kinetic parameters of the model are K_c, K_o, k_p and V_{cmax} are assumed to follow the Arrhenius-type relationship, in an analogous manner to the implementation of the FvC model described in Appendix B, with no change in activation energy over the temperature range examined. The values of these parameters are therefore expressed relative to their values at T_{calib}°C , $K_{c(Tcalib)}$, $K_{o(Tcalib)}$, $k_{p(Tcalib)}$ and $V_{cmax(Tcalib)}$ respectively using equations of the form of B.13, B.14 and B.15.

8.3.7. Plant Respiration

Canopy flux measuring techniques such as eddy covariance, measure the net ecosystem flux (NEF). NEF is the small difference between large CO_2 influx term resulting from photosynthesis and a large CO_2 efflux term resulting from respiration. Respiration is composed of above ground leaf and stem respiration and below ground respiration resulting from root respiration and microbial activity. For many ecosystems the below ground component (commonly referred to simply as soil respiration) accounts for a least half of the total CO_2 efflux (Norman, Garcia and Verma 1992).

Above Ground Respiration

Plant respiration is modelled as an empirical, exponential function of temperature, specified by a temperature coefficient. The response of respiration to nitrogen is assumed to be linear, specified by the slope and intercept of this relationship. The combined response can be described by:

$$R = \frac{\left(R^N_{slope} \cdot N + R^N_{\text{int}\,ercept}\right)e^{K_r T_{leaf}}}{e^{K_r T_{cal}}} \quad (8.20)$$

Where: R^N_{slope} is the slope of the respiration *vs* nitrogen relationship

$R^N_{intercept}$ is the intercept of the respiration *vs* nitrogen relationship

N is the leaf nitrogen concentration

K_r is the coefficient relating respiration to temperature

T_{leaf} is the leaf temperature

T_{cal} is the temperature at which the respiration *vs* nitrogen relationship was established.

Below Ground Respiration

Soil respiration, R' is modelled as a simple, empirical, exponential function of temperature in an analogous manner to above ground respiration, except that a nitrogen dependence term is not included:

$$R' = R'_0 e^{(k_{soil} T_{sol})} \tag{8.21}$$

Where: R'_0 is the soil respiration rate at 0 °C

k_{soil} is the soil respiration temperature coefficient

T_{soil} is the soil temperature

8.3.8. Light Use Efficiency

Light use efficiency (LUE) models of vegetation production (Monteith 1977) are discussed in Chapter 5. In such models biomass production is estimated directly from the amount of radiation absorbed by the vegetation, through a light use efficiency factor ε. There is some confusion in the literature surrounding ε since its value depends upon whether absorbed or intercepted radiation is used and whether both above and below ground or just above ground biomass production is considered. Consequently, two default methods are included in FRAMEWORK, as follows:

$$P_p = \varepsilon Q_a \tag{8.22}$$

$$P_{ag} = \varepsilon_{ag} Q_a \tag{8.23}$$

Where: P_p is the total above and below ground net primary production

ε is the light use efficiency for above and below ground biomass production

P_{ag} is the above ground net primary production.

ε_{ag} is the light use efficiency for above ground biomass production only

Q_a is the absorbed photosynthetically active radiation (PAR)

The exact configuration of FRAMEWORK using the LUE approach depends upon what factors are considered to impose a limitation on ε. For example, if it is required to study the potential effects of environmental and soil water constraints upon ε then FRAMEWORK might be configured to consist of the following objects **environment**, **radiation**, **soil water** and **light-use-efficiency**.

8.3.9. Other methods

Radiation Partitioning

The incident radiation flux, and fraction of diffuse radiation in both PAR and NIR wavelengths, is calculated from incident global radiation using the analysis of Weiss and Norman (1985): an empirical technique requiring only the solar zenith angle in addition to the incident global radiation as inputs.

Leaf Optics

The PROSPECT model (Chapter 4) is used to calculate leaf optical properties.

Aerodynamic Conductance

For expediency, two models for the aerodynamic conductance (g_b) are used depending on the canopy type and roughness and model configuration (big-leaf or multi-component):

For big-leaf canopies g_b is given by:

$$g_b = \frac{k^2 u(z)}{\ln\left(\frac{z-d}{z_0}\right)^2} \tag{8.24}$$

Where k is the von Karman constant (0.41)

z is the height at which the wind speed was measured

z_0 is the roughness length

d is the zero plane displacement

$u(z)$ is the wind speed

The default values for the roughness height and zero plane displacement are: $z_0 = 0.1 \times$ canopy height; $d = 0.7 \times$ canopy height.

For multi-component canopies the aerodynamic conductance is required for each component which may be at different heights within the canopy. Consequently, it is necessary to know the wind speed profile within the canopy. For this, the formulation of Landsberg and James (1971) is used, viz.:

$$u(z) = u_h \left[1 + \alpha \left(1 - \frac{z}{h} \right) \right]^{-2} \tag{8.25}$$

Where: u is the wind speed at height z

u_h is the wind speed at height h

α is an empirical coefficient and is a function of the vegetation drag coefficient, the leaf area density the canopy height and the eddy viscosity.

z is the vertical height

h is the canopy height

α is given by:

$$\alpha = h \frac{\sqrt{C_d u_h a}}{6 K_m} \tag{8.26}$$

Where: C_d is the foliage drag coefficient

a is the leaf area density

K_m is the eddy viscosity ($m^2 s^{-1}$)

Various empirical forms for α have been used in studies; Landsberg and Jarvis (1971) derived the following equation from studies of Sitka spruce canopies:

$$\alpha(z) = 6.85 \sqrt{\frac{a(z)}{u_h}} \tag{8.27}$$

8.4. THE HAPEX-SAHEL EXPERIMENT

HAPEX-Sahel (Hydrological and Atmospheric Pilot Experiment in the Sahel) was an international land-surface-atmosphere observation programme undertaken in western Niger in 1992 (Goutorbe *et al.* 1994). The intensive field campaigns obtained measurements of atmospheric, surface and sub-surface processes in a one degree square north-east of the capital city Niamey. Sahelian vegetation consists of annual grasses and scattered bush steppe toward the north with perennial grasses, scattered trees, and extensive rain-fed cultivation in the south (White 1983). The vegetation is strongly seasonal and virtually all woody species are deciduous, all herbs are either annual or die back to the ground

each dry season. In the north, herbs are green for approximately one month, increasing to three months in the south. Three supersites (Southern SSS, Central western CWSS and Central Eastern CESS) were defined inside the one degree x one degree square that composed the HAPEX field site, each approximately 20 x 20 km. The southern and central west sites were intended primarily for surface flux and energy balance studies and this demanded uniform stands large enough to measure the fluxes associated with the surface type within the stand, without significant influences from surrounding surface types.

The HAPEX-Sahel experiment was chosen as a source of data for validating the validation-optimization approach for integrating remotely sensed data into the FRAMEWORK, described in Chapter 5. The unique data requirements for validating the modelling approaches used here make finding data sources difficult. Data from HAPEX-Sahel are readily available both from an on-line database, and in the form of CD-ROMS, and include most of the necessary data e.g. the temporal variation of canopy structural properties and biomass, absorbed radiation, leaf and canopy scale flux measurements with associated meteorological data, leaf and soil optical properties, vegetation index estimates and canopy BRDF data. Although data were measured for a number of different species, the most comprehensive data were collected for millet (*Pennisetum glaucum*), consequently millet was chosen as the subject of this study. Millet is a C_4 cereal grown as a subsistence crop in areas where rainfall is inadequate for other cereals.

Mean annual rainfall at Niamey is about 560 mm, but the last 25 years have been marked with persistent drought and the mean for this period is 495 mm. The average date of beginning of rains at Niamey is 12 June and the average length of the growing season is 94 days (Sivakumar 1989). Mean minimum and maximum temperatures for the rainy season are 22 and 34 °C, respectively, increasing northwards (Sivakumar 1989). The rainfall during the field campaign in 1992 was characterized by an early start of the rains followed by a dry spell from early June until mid July, with rainfall above average in August and early September. The rainfall total at the SSS was 787 mm, whereas at the central east site it was only 410 mm, however, over the whole study area rainfall was 537 mm. Solar radiation varied somewhat in the dry season, possibly the result of atmospheric dust. Air temperature rose with the increase in solar radiation but then declined during the rainy season, with the shift of energy transfer from sensible heat to latent heat. Night-time temperatures were low in the dry season, but increased during the growing season to approximately 25 °C. Daytime temperatures reached a maximum of 42 °C before the rains started, but fell to about 32 °C in the mid rainy season. The soil in the region is extremely sandy, composed of approximately 95% sand with 0.15% organic carbon by mass. The water table was at a depth of 25 m.

Vegetation development in 1992 was reported as being close to the average, but there was a marked delay in the start of the growing season. The individual supersites differed quite markedly in their vegetation development as indicated by their NDVI. The NDVI in the southern site increased from June and remained above 0.2 for three months. At CESS, NDVI increased one month later and at CWSS a month later again, in August. In both central sites the NDVI was above 0.2 for only two months. Millet was planted between 16 May and 5 June at the SSS, with an average density of 4600 ±1260 pockets per hectare, corresponding to a spacing of approximately 1.4 m, with 3 plants per pocket after thinning. Senescence commenced in late August and harvesting in early September produced a yield of 650 ± 390 kg grain ha^{-1}. The dry period in June caused the millet crop in most fields sown in May to fail except in the south west. Fields elsewhere were re-sown, sometimes more than once. In the SSS, millet was harvested between 10 and 15 September but not until about 10 October in the central sites.

8.5. PARAMETERIZING FRAMEWORK FOR MILLET

The following sections describe the parameteriztion of FRAMEWORK for modelling millet canopy fluxes and growth during the 1992 growing season in Niger. The measured data used were obtained from the Hapex-Sahel database CD-ROM (HAPEX-Sahel Information System CD-ROM 3: Ground data v3 08/95). Additional millet data were kindly provided by Peter Levy, University of Edinburgh. These included, stomatal conductance measurements, leaf-scale flux measurements and respiration data. Additional NDVI and directional reflectance data were supplied by Agnes Bégué.

The millet canopy was represented by a single-layer, big-leaf model. Light interception was calculated using the Bégué model (Bégué 1992), coupled with the PROSPECT model of leaf optical properties. Photosynthesis was modelled using a modified version of the Collatz C_4 photosynthesis model (Collatz, Ribas-Carbo and Berry 1992). An original Fortran 77 version of the Collatz model was generously supplied by Jon Lloyd, Australian National University, Canberra. Respiration was modelled according to a simple empirical exponential relationship with temperature. All other methods used were the defaults, described in section 8.3, suitably parameterized for millet using measured data.

8.5.1. Radiation Interception

The radiation intercepted, reflected and absorbed by a vegetation canopy depends strongly upon the structure and reflective properties of the canopy and the incident radiation regime. Thus, the key variables are: leaf area index, leaf angle distribution, degree of clumping of branches/foliage, leaf and soil optical properties, solar zenith and azimuth angles and the diffuse fraction of incident radiation. Many models of the radiation transfer of plant canopies have been developed for homogeneous canopies, e.g. the SAIL model, Chapter 3, but few studies have considered spatially discontinuous

canopies. Because structure plays such an important rôle in determining the radiation environment of a vegetation canopy (Monteith, 1981), one or two-dimensional radiative transfer models are unsuitable for modelling sparse canopies such as those of millet.

Radiation interception or absorption is frequently derived from NDVI data, however this relationship is subject to variability because of changes in sun-sensor geometry and background reflectance (Choudhury, 1987). These effects are amplified for sparse canopies because of the small canopy coverage. Bégué (1992) developed a geometric model of radiative transfer in sparse canopies where vegetation clumps are represented as porous cylinders (Chapter 3). The model is split into two submodels describing the macro and microstructure of the canopy. The macrostructure submodel calculates the directional interception efficiency as a function of the height, porosity and spacing of the cylinders. The microstructure submodel uses average transmittance theory to compute the directional porosity from:

$$p(\theta_s) = \exp[-K_c(\theta_s, \eta) H L_v (1 - \tau_F)] \tag{8.28}$$

Where: p is the porosity

θ_s is the solar zenith angle

H is the cylinder height

η is the shape coefficient equal to the ratio of the cylinder radius (R) and height (H)

K_c is the extinction coefficient of the clump

L_v is the leaf area density

τ_F is the leaf transmittance

Figure 8.10 shows the measured daily average porosity for millet. It can be seen that the porosity is almost constant through the growing season with a mean value of 0.56 and a standard deviation of 0.05. This initially surprising result can be explained with reference to Figure 8.33. As the plant grows the LAI, canopy and height and the radius of the clumps all increase in such a way as to maintain a nearly constant leaf area density, and consequently a near constant porosity. This contrasts with homogeneous vegetation canopies for which the porosity decreases with growth.

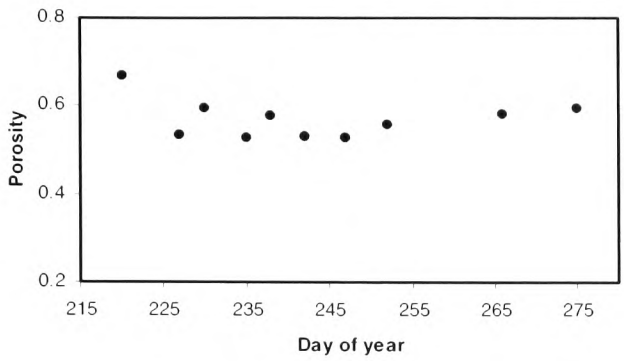


Figure 8.10 Porosity of millet at SSS during the 1992 growing season, measured by A. Bégué (HAPEX-Sahel database).

Leaf angle distribution measurements for simulations using the Bégué model were kindly provided by P. Lewis, University College London, and are shown in Figure 8.11

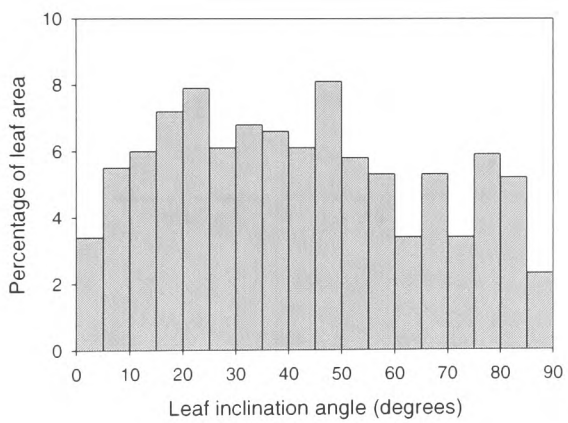


Figure 8.11 Millet leaf angle distribution during August at SSS, measured by P. Lewis, University College London.

In order to model canopy light interception and absorption in different wavebands, the Bégué model was coupled to the PROSPECT model of leaf optical properties (Chapter 4). The good fit between PROSPECT modelled leaf reflectance and transmittance, and measured values for healthy millet leaves millet, is illustrated in Figure 4.11a,b. The PROSPECT model parameters for the modelled optical properties shown in Figure 4.11a,b were obtained by simulated annealing optimization: the values for the parameters are shown in Table 8.2.

Figure 8.12 shows the comparison between the measured PAR interception for millet, and the interception modelled by SAIL. Since the SAIL model assumes a homogeneous canopy, two values of LAI were used to parameterize the model, the total LAI of the plot, and the LAI within the individual clumps. For the latter case, the canopy interception estimates were weighted by the fractional vegetation cover of the plot. Using the LAI of the plot resulted in a significant overestimation of interception as a consequence of very high values of interception being predicted near midday when solar radiation was maximal. Using the LAI of the clumps resulted in near-constant interception estimation which significantly underestimated the measured value. This was a consequence of the LAI within the clumps being around 5, resulting in almost compete interception.

Clearly the SAIL model is not suitable for modelling the PAR interception of a sparse canopy such as millet. However, the Bégué model, which was specifically designed for sparse canopies, was able to model PAR interception much better. Figure 8.13 shows the comparison between measured interception and interception modelled using the Bégué model. The fit is excellent during the growing season ($r^2 = 0.91$). The fit during the senescence period is poorer but still quite good ($r^2 = 0.84$). This probably results from larger parameter error during this period (P. Levy *pers. comm.*) The Bégué-modelled PAR absorption, shown in Figure 8.14, also fits the measured data well with a slight tendency to overestimation at low absorptance, probably caused by variation in the soil reflectance values (P. Lewis *pers. comm.*).

Parameter	*Value*
N	1.61
C_{ab}	50.8 μg cm^{-2}
C_w	0.0005 cm
C_p	0.001 g cm^{-2}
C_c	0.004 g cm^{-2}

Table 8.2 PROSPECT model parameters fitted to measured millet reflectance and transmittance data at the CWSS using simulated annealing optimization. Measurements made by N. Hanan (HAPEX-Sahel database)

Figure 8.15 shows a comparison between the modelled and measured hourly PAR interception averaged over the period from day 231 to day 236 in 1992. With the exception of a spurious point caused by measurement error, the model results fit the data extremely well, providing further evidence that the Bégué model represent the PAR interception of millet accurately.

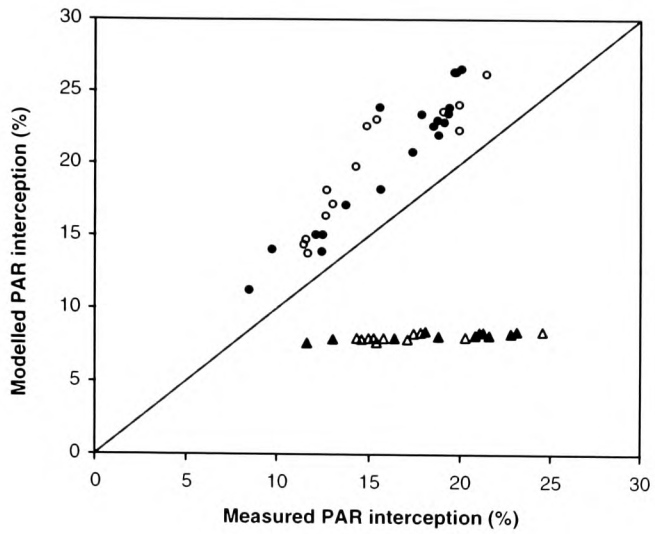


Figure 8.12 Comparison between measured millet PAR interception and interception modelled using SAIL with the LAI of the canopy, for the growing season (●) and senescence period (o); and the LAI of the clumps, for growing (▲) and senescence (Δ) periods. (CWSS. Data measured by A. Bégué obtained from the HAPEX-Sahel database.)

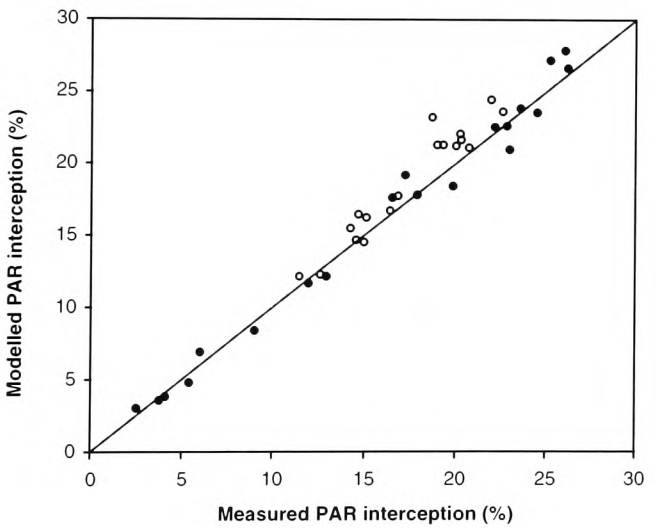


Figure 8.13 Comparison between measured millet PAR interception and interception modelled using the Bégué for the growing season (●) and senescence period (o). (CWSS. Data measured by A. Bégué obtained from the HAPEX-Sahel database.)

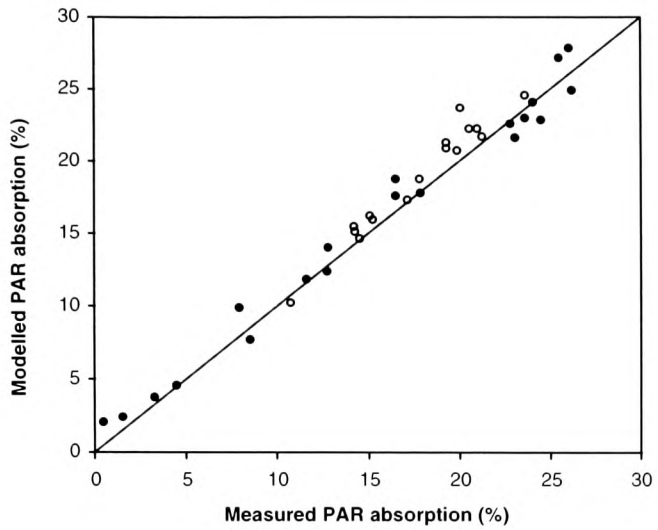


Figure 8.14 Comparison between measured millet PAR absorption and absorption modelled using the Bégué model for the growing season (●) and senescence period (o). (CWSS. Data measured by A. Bégué obtained from the HAPEX-Sahel database.)

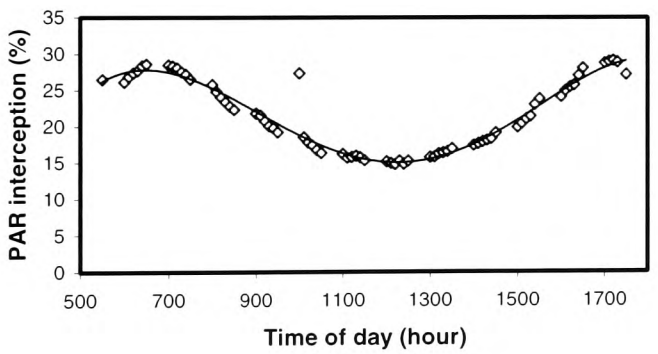


Figure 8.15 Hourly PAR interception for millet averaged over a five day period from day 231 to day 236, during the 1992 growing season (points) compared to PAR interception estimated using the Bégué model (line). (CWSS. Data measured by A. Bégué, obtained from the HAPEX-Sahel database.)

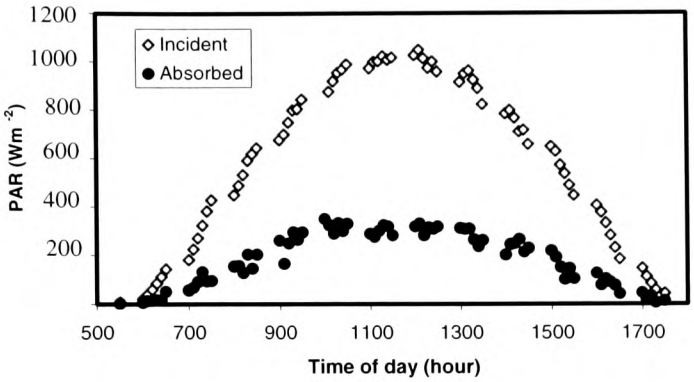


Figure 8.16 Hourly incident PAR averaged over a five day period from day 231 to day 236, during the 1992 growing season compared to PAR absorption by millet estimated using the Bégué model.

8.5.2. Canopy Reflectance and NDVI Prediction

The ability of the Bégué model to simulate millet canopy directional reflectances and NDVI values was also investigated. Soil reflectance properties exert a strong influence on surface reflectance and NDVI values (Chapter 3). This effect is enhanced for sparse canopies, such as millet, over denser canopies, because a larger proportion of the soil is visible. The SOILSPECT+ model was used to simulate millet soil background reflectance during the growing season. The model was fitted to data by simulated annealing inversion. The results of this inversion are illustrated in Figure 8.17, which shows the comparison between the measured soil spectral reflectance and that modelled using SOILSPECT+. Figure 4.19 shows the comparison for average soil reflectances measured over the growing season. In both cases the model fit is good, indicating that the SOILSPECT+ model is able to represent the measured soil reflectances well.

Canopy reflectance is also a function of leaf optical properties. The Bégué model was coupled with the PROSPECT model (Chapter 4) within FRAMEWORK; this model will hereafter be referred to as simply the Bégué model. Figures 4.11 and 4.12 compare PROSPECT modelled millet leaf reflectance and transmittance with measured data. These Figures shown that PROSPECT models the optical properties of millet leaves in a variety of stages of health very well, with the exception of chloritic and senescent leaves.

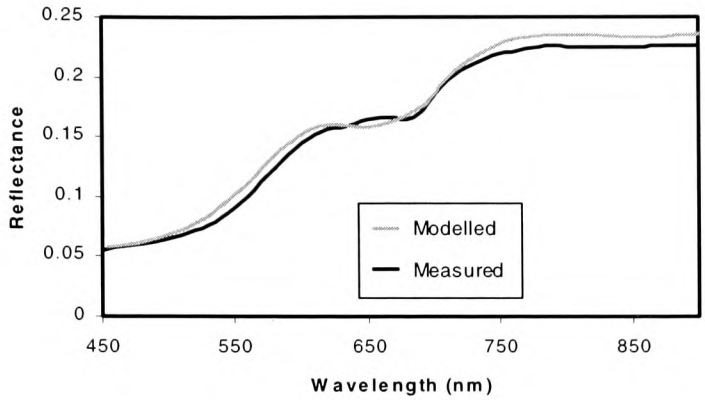


Figure 8.17 Comparison of soil reflectance modelled using the SOILSPECT+ model fitted to measurewd data using simulated annealing optimization, and soil reflectance measured for the HAPEX-Sahel site. (Data obtained from the HAPEX-Sahel database).

Initial modelling results showed a significant scatter between modelled and measured reflectances stemming from predictions made near the end of the growing season, after senescence had started. At this stage of development, the bottom, background layer, composed of a mixture bare soil and senescent leaves, leads to errors in prediction of the background reflectance. Therefore, a modification was introduced to compute the background reflectance as a linear combination of soil and senescent leaf reflectance, with the proportion of senescent leaf reflectances assumed to increase linearly from 0% at maximum crop LAI to 100% at full senescence. Figure 8.18, Figure 8.19 and Figure 8.20 show a comparison of the measured millet directional reflectances in red green and NIR wavebands with predictions made by the Bégué model for a variety of growth stages. The model fit is extremely good with root mean square errors of 1.5%, 2.1% and 4.1% for green, red and NIR, respectively.

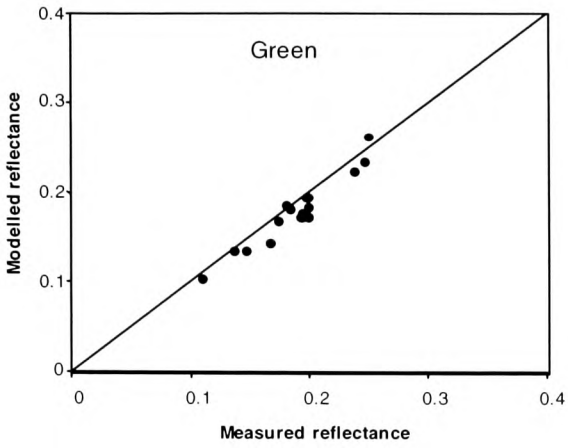


Figure 8.18 Directional reflectance measured for millet and modelled using the Bégué model in the green part of the visible spectrum. The model was parameterized using data sources that were independent of the canopy-scale data measured by A. Huete shown in the Figure. CWSS. Data obtained from the HAPEX-Sahel database.

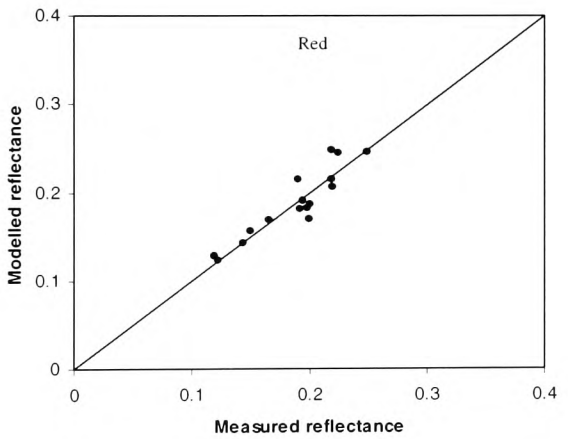


Figure 8.19 Directional reflectance measured for millet and modelled using the Bégué model in the red part of the visible spectrum. The model was parameterized using data sources that were independent of the canopy-scale data measured by A. Huete shown in the Figure. CWSS. Data obtained from the HAPEX-Sahel database.

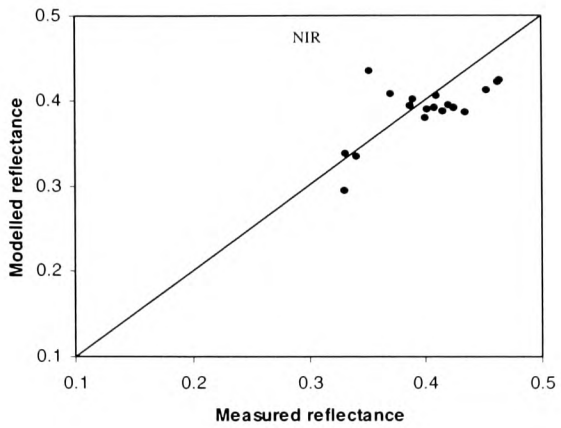


Figure 8.20 Directional reflectance measured for millet and modelled using the Bégué model in the near infrared part of the spectrum. The model was parameterized using data sources that were independent of the canopy-scale data measured by A. Huete shown in the Figure. CWSS. Data obtained from the HAPEX-Sahel database.

Figure 8.20 illustrates that there is more scatter for NIR than for red and green reflectances. Most of this scatter can be explained by the low contrast between canopy and soil reflectances in this waveband. The Bégué model assumes single scattering, which although a good approximation for visible wavebands, does not allow for the scattering characteristics usually observed at infrared wavelengths. However, for sparse canopies the assumption of single scattering in the NIR may be a valid approximation because of the high probability of scattered light escaping from the canopy. This is borne out by the quite small RMS error for the modelled NIR reflectances at 4.1%, despite the low contrast between leaf and soil reflectances.

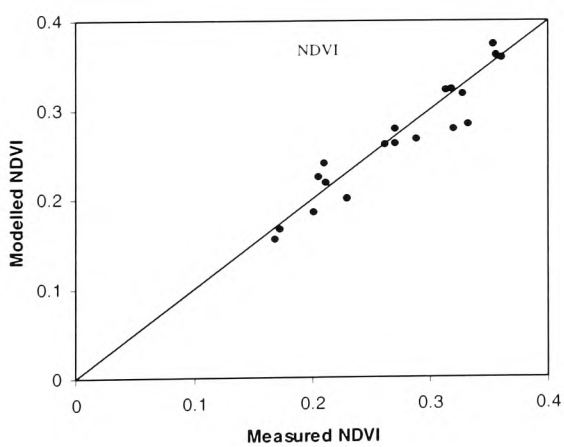


Figure 8.21 Millet canopy NDVI measured for millet and modelled using the Bégué model. The model was parameterized using data sources that were independent of the canopy-scale data measured by A. Huete shown in the Figure. CWSS. Data obtained from the HAPEX-Sahel database.

Figure 8.21 shows the comparison between measured NDVI values for millet and NDVI modelled using the coupled Bégué and SOILSPECT+ models. The predicted NDVI values agree well with the measured data with a RMS. error of 0.042, indicating that the model is capable of simulating NDVI values for millet canopies to a high degree of accuracy.

8.5.3. Millet stomatal Conductance

The stomatal conductance model was fitted to the measurements of the response of stomatal conductance to incident light and saturation vapour pressure deficit, kindly provided by Peter Levy, University of Edinburgh, using a trained genetic algorithm approach, Table 8.3. The response of g_s to PPFD predicted by the model using the parameters in Table 8.3 is shown in Figure 8.22.

Parameter	*Symbol*	*Fitted value*
The maximum allowed value for g_s	g_{smax}	0.771 mol m^{-2} s^{-1}
The stomatal conductance in the dark	g_{sdark}	0.183 mol m^{-2} s^{-1}
The initial slope of the g_s *vs* light curve	δ_{g_s}	0.9 x10^{-3} mol m^{-2} s^{-1} mol^{-1} quanta
The value T of g_s is a maximum	T_{opt}	30 °C
Temperature below which $g_s = 0$	T_{low}	-5 °C
Temperature above which $g_s = 0$	T_{high}	40 °C
Minimum value for g_s as a function of water	Ψ_{min}	2.5 x10^6 Pa
Threshold for the g_s *vs* soil water curve	$\Psi_{threshold}$	0.4 x10^6 Pa

Table 8.3 Parameters fitted to the FRAMEWORK stomatal conductance model (Jarvis-type) for millet, using data measured by P. Levy, University of Edinburgh at the SSS. Measurements made using an ADC Leaf Chamber Analysis system (LCA3) and a PLC3(N) ADC Parkinson Leaf Chamber during February, March, June and October 1992.

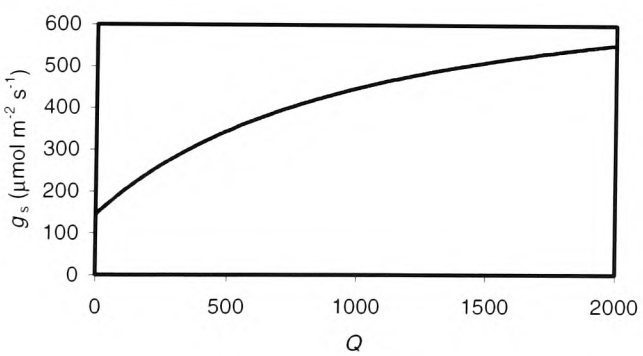


Figure 8.22 Response of stomatal conductance (g_s) to absorbed PPFD (Q) predicted by the stomatal conductance model using the parameters given in Table 8.3.

A considerable amount of scatter was present in the measured data, consequently the model fit was relatively poor, $r^2 = 0.31$ with variations in PPFD and VPD accounting for only 31% of the variation

in g_s. However, no clear relationship could be established between stomatal conductance and either leaf or air temperature, and no data were available on the variation of g_s with soil moisture. The residual variation is attributed to inter-shoot variability, as much better fits can be achieved using data from individual shoots (P. Levy, *pers. comm.*). Although observations showed considerable variability, the fitted model was able to reproduce the average diurnal trend of stomatal conductance reasonably well (Figure 8.23).

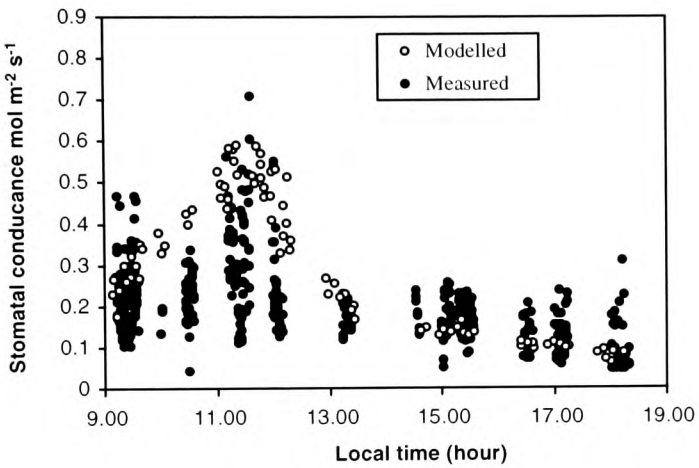


Figure 8.23 Comparison of the diurnal trend in stomatal conductance modelled using the fitted Jarvis-type model and measured for millet at the SSS on 22/9/1992.

8.5.4. Millet Respiration

Above Ground Respiration

Respiration was modelled using an empirical, exponential temperature response. Since the canopy is modelled using the big-leaf approach, and since data for canopy nitrogen content and distribution were not available, a nitrogen response function for respiration was not used in the model. Instead the simple empirical temperature response for respiration, of the form, $R = R_0 \exp(K_r T)$, where R_0 is the respiration a 0 °C and K_r is the coefficient relating respiration to temperature, T, was fitted to data measured by P. Levy, University of Edinburgh, as shown in Figure 8.24. Only a limited number of respiration measurements were available, and these exhibit a considerable amount of scatter; however, an increase in respiration with temperature is detectable. The parameters of the fitted model are shown in Table 8.4. As can been seen from the r^2 values of 0.38, temperature effects explain 38% of the variation in the data. The modelled and measured respiration values by the model are quite large relative to other species at the HAPEX site, exceeding 4 μmol m^{-2} s^{-1} at 30 °C. This is probably a result of the high nutrient status of the fertilized crop relative to the unfertilized shrub savannah. Table 8.4 also shows the Q_{10} value for the respiration model [$Q_{10} = \exp(10K_r)$] which is within the range of values usually observed in the field of between 1.4 and 2.2.

Parameter	*Value*
Respiration rate at 0 °C, R_0	0.697×10^{-6} mol m^{-2} s^{-1}
Respiration temperature coefficient, K_r	0.047
r^2	0.38
Q_{10}	1.59

Table 8.4 The parameters of the respiration model fitted to measured data using simulated annealing optimization.

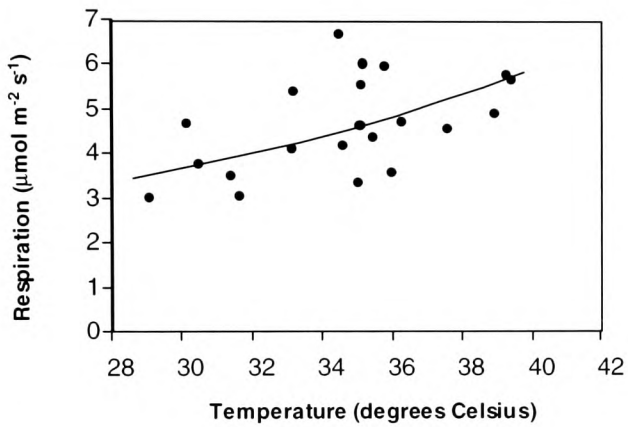


Figure 8.24 The relationship between leaf dark respiration and leaf temperature for millet at the SSS. Points are measurements made by P. Levy, University of Edinburgh, throughout the growing season in 1992. The range of temperatures was obtained by making measurements at different times of day. The line shows the fit of the respiration model.

Below Ground Respiration

Soil CO_2 efflux measurements were generously provided by P. Levy, University of Edinburgh. Figure 8.25a shows the diurnal trend in soil respiration and soil temperature, measured at the fallow site from 27 to 29 June. Although a response of soil CO_2 efflux to temperature is apparent in the Figure, the response lags behind the diurnal course of soil temperature at 1cm by about four hours. This suggests that the main source of respiration is much deeper in the soil than 1cm and is therefore probably associated with the roots of the bushes (ground flora had not developed when the measurements were made). There is a paucity of data and consequently a large degree of uncertainty associated with the soil respiration estimates. However, the small size of the fluxes, less than 0.18 µmol m^{-2} s^{-1}, suggests that any error may not be too significant. It is important to note, however, that soil respiration fluxes may be larger later in the growing season when vegetation had developed further.

Figure 8.25b shows the relationship between soil respiration and proximity to vegetation at the fallow site. There is a significant decrease in soil respiration with distance from the nearest bush. This

spatial heterogeneity may be related to root density and/or litter content of the soil. As described in Section 8.4, the soil in the region is extremely sandy, composed of approximately 95% sand with 0.15% organic carbon by mass. Consequently, the spatial pattern of soil respiration reinforces the assumption that the source of respiration is associated with the roots of the bushes, since there is little organic matter to support microbial activity.

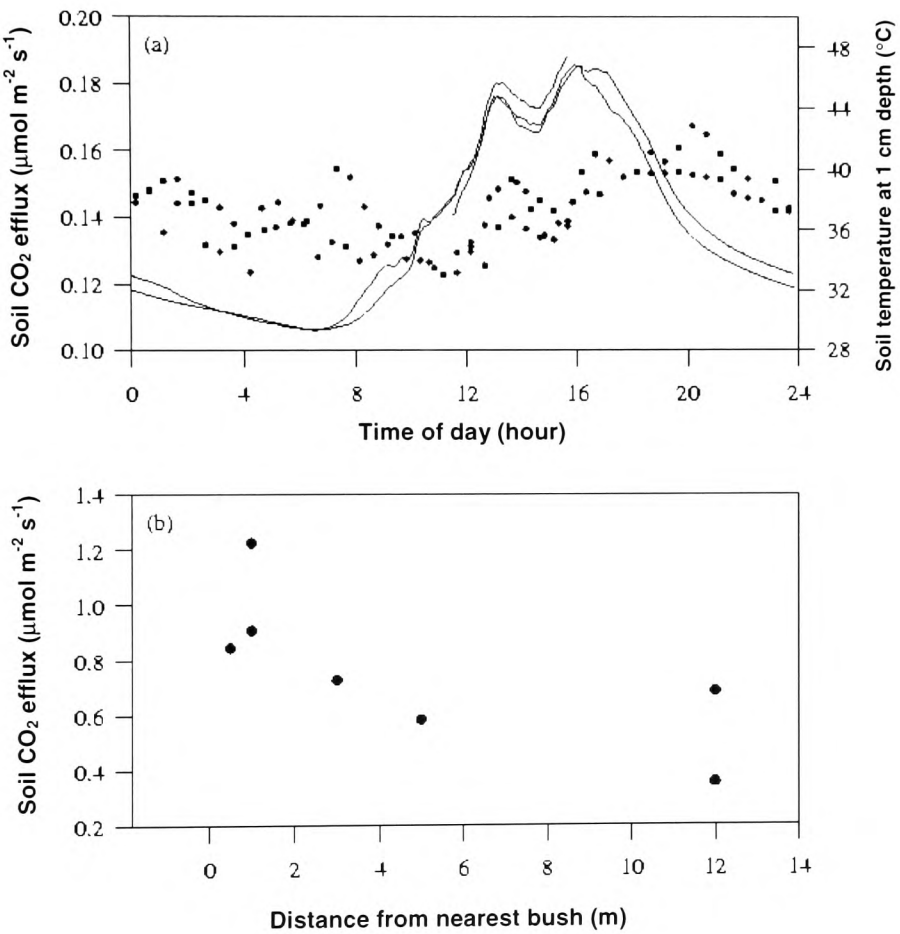


Figure 8.25 (a) Relationship between soil respiration (symbols) and soil temperature (lines) measured at the fallow site between 27 and 29 June 1992. (b) Relationship between soil respiration and proximity to bushes, measured along two transects at the fallow site on 15 June 1992. The soil temperature range for these measurements was 28 to 34 °C. All measurements made by P. Levy, University of Edinburgh (*pers. comm.*)

Unfortunately, no soil CO_2 efflux measurements were available for the millet site. However, root biomass measurements were available (J. Brouwer, HAPEX-Sahel database) and these were used to estimate root respiration by using the parameters derived from measurements of leaf respiration shown in Table 8.4, and assuming that the respiration rate per unit mass of roots was the same as that for leaves. The millet soil respiration calculated using this technique is shown in Figure 8.26.

Although there is no evidence that the technique used for deriving millet soil CO_2 efflux is valid, this approach seems more justifiable than basing predictions on the small number of soil respiration measurements shown in Figure 8.25.

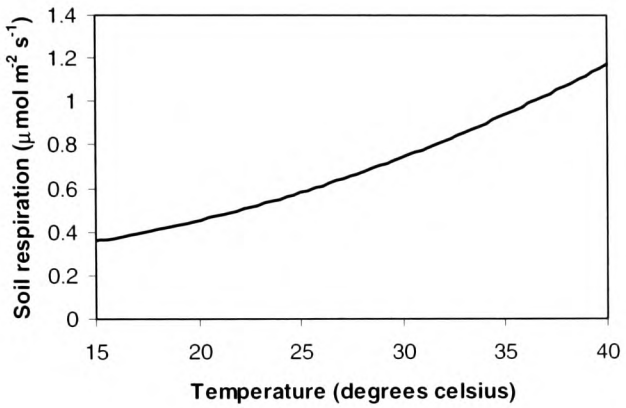


Figure 8.26 Soil respiration for millet at the SSS. CO_2 efflux resulting from microbial activity was assumed to be negligible, therefore the soil respiration is composed entirely of root respiration. Root respiration was calculated using the parameters derived from measurements of leaf respiration shown in Table 8.4, and assuming that the respiration rate per unit mass of roots was the same as that for leaves. A value of 0.24 was used for the ratio of above ground biomass per uit area to below ground biomass per unit area and was assumed to be constant throughout the growing season. This value was derived from measurements of root biomass made at the SSS by J. Brouwer throughput the growing season (data obtained from HAPEX-Sahel database).

8.5.5. Millet Photosynthesis

The Collatz C_4 photosynthesis model (Collatz, Ribas-Carbo and Berry 1992) was used to model carbon assimilation by millet. For simplicity, this model will be known hereafter as the CRCB model. Although this model, as described by the authors, is a coupled photosynthesis-stomatal conductance model, using the Ball-Berry model of stomatal conductance, in this study the model has been modified in order to incorporate it into the FRAMEWORK model structure, and a Jarvis-type model is used for stomatal conductance.

C_4 and C_3 photosynthesis differ in several important biochemical and physiological properties. In C_3 photosynthesis CO_2 fixed by rubisco is obtained directly from the intercellular air spaces by diffusion. In C_4 species, CO_2 to delivered to rubisco by a metabolic pump that concentrates CO_2 in the bundle sheath cells, which in turn inhibits photorespiration. These differences in function between C_3 and C_4 photosynthesis result in differing sensitivities to environmental factors such as CO_2 and O_2 concentration, temperature and PPFD.

Table 8.5 shows the parameters of the Collatz *et al* photosynthesis model that were used in the millet simulations. Parameters were obtained either from literature values, experimental data generously provided by J. Lloyd, Australian National University, Canberra, or by fitting the model against leaf and leaf scale flux measurements made by P. Levy, University of Edinburgh at the SSS during the 1992 growing season.

Parameter	Symbol	Value	
M-M constant for oxygenation	$K_{o(Tcalib)}$	140 Pa	1
M-M constant for carboxylation	$K_{c(Tcalib)}$	34000 Pa	
Intercellular partial pressure of O_2	O_i	21.3 x10^3 Pa	
Quantum yield for rubisco	α_r	0.11 mol mol^{-1}	
Diffusitivity of O_2 relative to CO_2	D	40	
Quantum yield for PEP	α_p	0.167 mol mol^{-1}	
Specificity of rubisco for CO_2 relative to O_2	τ	2600	
PEP rate constant for CO_2	k_p	0.7 mol m^{-2} s^{-1}	2
Activation energy for K_o	E_o	59.3x10^3 J mol^{-1}	
Activation energy for K_c	E_c	36x10^3 J mol^{-1}	
Activation energy for V_{cmax}	E_v	42x10^3 J mol^{-1}	
Activation energy for k_p	E_k	42x10^3 J mol^{-1}	
Max. carboxylation rate for rubisco	$V_{camx(Tcalib)}$	20x10^{-6} mol m^{-2} s^{-1}	
Calibration temperature	T_{calib}	30	
Intercellular resistance to CO_2 diffusion	r_c	712.23 m s mol^{-1}	3
Proportion of O_2 production in bundle sheath	λ	0.475	
Fractional rubisco quantum requirement	f	0.652	

Table 8.5 Parameters fitted to the FRAMEWORK photosynthesis model (Collatz *et al* 1992) for millet. Parameters derived from (1) Collatz *et al* (1992), (2) J. Lloyd, Australian National University, Canberra (Pers. Comm.), (3) fit of model to measured leaf and canopy CO_2 fluxes shown in Figure 8.24, Figure 8.27 and Figure 8.28.

Photosynthesis Model Results

Figure 8.27 shows the relationship between millet leaf CO_2 flux, expressed on a leaf-area basis, and incident PPFD measured using a leaf chamber and modelled using the CRCB model within the FRAMWORK model using the parameters shown in Table 8.1. Measured data were collected by P. Levy, University of Edinburgh. Figure 8.28 shows the same relationship for canopy scale fluxes, expressed on a ground area basis, measured using eddy covariance techniques and FRAMEWORK model predictions. Canopy scale flux data were collected by a team from the University of Edinburgh led by J. Moncrieff (Moncrieff *et al.* 1997) and were independent of the leaf scale mesurements used to derive some of the parameters in Table 8.5. There is rather more scatter amongst the canopy scale measurements that probably reflects day to day variation in wind direction. The wind direction determines which area of the crop lies within the fetch of the sensor, so that the scatter in the signal may be the result of different areas of the crop with different growth rates on different days.

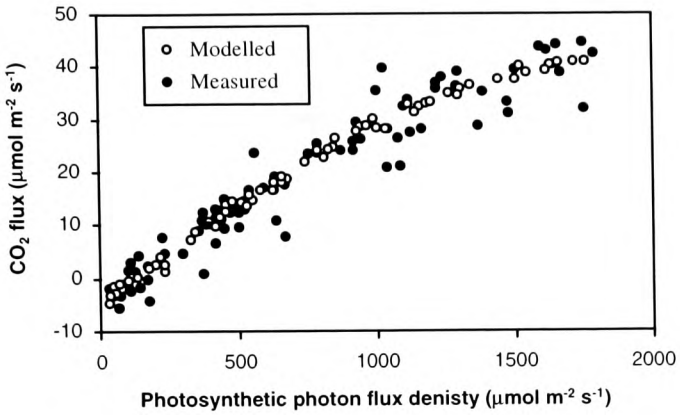


Figure 8.27 The relationship between modelled and measured millet leaf CO_2 flux and incident PPFD at the SSS. Measured fluxes are data from leaf chamber measurements by P. Levy, University of Edinburgh, collected throughout July and August 1992. Shading of the leaf chamber was used to manipulate PPFD. Modelled fluxes were predicted using the FRAMEWORK model

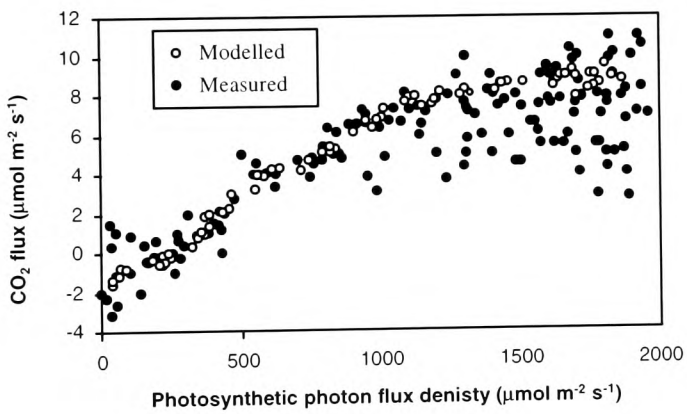


Figure 8.28 The relationship between modelled and measured millet canopy CO_2 flux and photosynthetic photon flux density (PPFD). Measured fluxes are hourly averaged eddy covariance data collected between 5th and 11th September 1992 at the SSS by (Moncrieff *et al.* 1997). LAI was measured as 0.46. Modelled fluxes were predicted using the FRAMEWORK model parameterized using leaf-scale data.

Canopy scale fluxes are significantly smaller in magnitude than leaf scale fluxes because the green leaf area index at the time of measurements was only 0.46 because the leaves were senescing just prior to harvest. In additional, although the leaf area is small, the individual plants, and the leaves on the plants, are clumped which resulted in a degree of mutual shading (see Section 8.5.1).

Both canopy and leaf scale fluxes show no saturating effect with increasing PPFD, typical of C_4 photosynthesis. The millet never becomes PPFD saturated for two reasons. The leaf conductance (g_s) rarely falls below 300 μmol m^{-1} s^{-1} in the cloud free conditions that frequently prevailed during the growing season. Moreover, even at small values of leaf conductance the C_4 metabolic pump maintains a supply of CO_2 to the carboxylation sites. Of the few points where g_s was small (less than 300 μmol m^{-1} s^{-1}), most were measurements from late in the season when leaves were becoming senescent. Ignoring these points, the light response is close to linear. Although temperature has an effect on respiration, it has only a small effect on C_4 photosynthesis in the range 27 to 43 °C, consequently PPFD is the only significant controlling variable.

8.6. Optimizing FRAMEWORK Growth Predictions

The development of millet over the whole growing season was modelled using FRAMEWORK, configured and parameterized as described in the preceding sections. The aim of this study was to investigate the potential for using remotely sensed data to validate or constrain the model as outlined in Chapter 5. Chapter 5 describes two approaches for coupling remotely sensed data to vegetation models, inversion-parameterization and validation-optimization. Numerous studies have adopted the inversion-parameterization approach and the application of this technique to the inversion of derivatives of the SAIL model is described in Chapter 7; however, little research into the use of the validation-optimization technique has been done. Therefore, this section will concentrate on the validation-optimization strategy.

The CRCB C_4 photosynthesis model requires a time-step of the order of 1 hour. It is not feasible to apply the validation-optimization approach to such models over the times scales (i.e. weeks) at which successive remote sensing scenes are typical available, using normal computational techniques. This is because stochastic optimization algorithms such as genetic algorithms or simulated annealing are very computationally intensive. In effect, these algorithms parameterize the model using random or pseudo-random sets of parameters and select the optimum set from a very large number of model runs. Consequently, to optimize an hourly time step model over a series of five remote sensing scenes each seven days apart, each model run with each parameter set would require $24 \times 7 \times 5 = 840$ iterations. Since each iteration consists of many floating point operations it is easy to see that the optimization would quickly become intractable using normal computing methods.

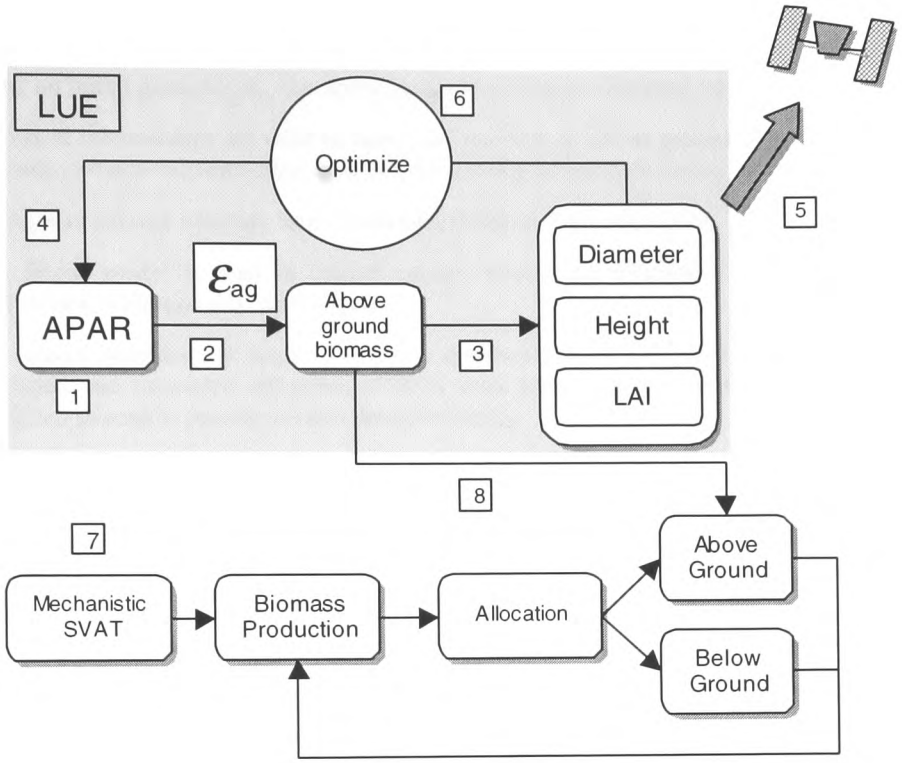


Figure 8.29 The validation-optimization approach applied to coupled light use efficiency and mechanistic SVATs with the FRAMEWORK model. The numbers refer to steps described in the text.

This Chapter describes two alternative approaches that are described below with reference to Figure 8.29: an *efficiency* strategy and an *allocation* strategy.

8.6.1. Efficiency Strategy

By configuring FRAMEWORK as a LUE model with a daily time step, the validation-optimization approach can be used to constrain the model output over time. In this approach, it is assumed that the relationships between above ground biomass production and LAI, canopy height and clump diameter are known. These relationships, established from data measured during the 1992 field campaign, are assumed to be invariant between individuals, and shown in Figure 8.36, Figure 8.37 and Figure 8.38. The relationships used by the model were obtained from interpolation on the measured data, fitted between zero abscissa and maximum on the abscissa, i.e. the senescent period was not considered. If relationships between above ground biomass and LAI, clump diameter and canopy height are known, then the value of the light use efficiency factor for above ground biomass production, ε_{ag}, can be found throughout the growing season. Consequently, the effects of environmental stress on ε_{ag} can be studied. This strategy shown in Figure 8.29 (shaded) proceeds as follows:

1. Starting from initial guesses for the canopy structural parameters, the Bégué model is used to calculate APAR.
2. Using an initial guess for ε_{ag}, the above ground biomass is estimated from the APAR.
3. Empirical relationships are used to relate any increase in above ground biomass to increases in canopy structural parameters such as LAI, canopy height and canopy diameter.
4. Changes in canopy structure have feedback effects on PAR absorption.
5. The Bégué model is used to predict canopy directional reflectance or NDVI values for comparison with remotely sensed data.
6. The model (defined by steps 1 to 5) is optimized to minimize the difference between predicted and measured reflectances/NDVI over time, thus constraining the trajectory of modelled structural parameters and determining ε_{ag}.

8.6.2. Allocation Strategy

The flexibility of the FRAMEWORK model enables complex mechanistic and simpler empirical SVATs each operating on different time-steps to be coupled together (Chapter 5). In this strategy, the coupling between the empirical and mechanistic sub-models is via biomass allocation. The concept is to use a constrained LUE model to obtain estimates of biomass allocation into either (i) different above ground components such as leaves and stems, or (ii) into broader categories of above and below ground components.

Steps 1 to 5 of the allocation strategy are the same as for the efficiency strategy, above. However, in contrast to the efficiency strategy, it is assumed that ε_{ag} is known in step 2. The increase in above ground biomass resulting from growth processes is calculated from the APAR using ε_{ag}. In step 6 instead of optimizing ε_{ag} using known allocation functions, the allocation functions themselves are optimized using. Above ground biomass is partitioned into components representing leaf, stem and spike using empirical relationships with the canopy structural variables LAI and canopy height. The allocation functions assume that the specific leaf area (leaf area per unit leaf biomass) is constant throughout the growing season and, therefore, that the total leaf biomass is a linear function of the leaf area index. Similarly, it is assumed that the stem biomass is a linear function of canopy height. The above ground biomass production and the above ground biomass allocation functions obtained from the optimization of the LUE model are used as constraints in the coupled mechanistic SVAT, so that steps 7 and 8 of the allocation strategy become:

7. The coupled mechanistic model predicts total carbon assimilation resulting from photosynthesis. Assimilate is allocated into above ground and below ground components.
8. The constrained trajectory of above ground biomass and the optimized biomass allocation functions are used as a feedback into the mechanistic model to constrain the biomass allocation function.

8.6.3. Measurements

The model parameterizations and results are complicated by the widely differing growth patterns and yields that occurred at the different supersites during the 1992 field campaign, Table 8.6. Data from the southern (SSS) and central western (CWSS) supersites were used for model simulations. Nine days of eddy covariance CO_2 flux measurements were available for the southern supersite as were a limited number of canopy structural data and NDVI measurements. Leaf scale photosynthesis respiration and stomatal conductance measurements were also available. At the CWSS detailed canopy structural data including biomass allocation were available at a much higher temporal frequency than at the SSS. Directional reflectance data were also available, however, no canopy or leaf scale flux measurements were made.

Parameter	*SSS*	*CWSS*
Density (clumps per hectare)	4600 ± 1260	3629 ± 585
Approximate sowing date	16/5	5/6
Max LAI	0.72	0.46
Date of max LAI	25/8	15/9
Harvest date	10/9 – 15/9	10/10
Yield	650 ± 390 kg ha^{-1}	

Table 8.6 Comparison of millet growth at the southern supersite (SSS) and the central western supersite (CWSS) during the 1992 growing season.

The main cause of the markedly different vegetation development at the individual supersites was a difference in rainfall pattern. The main feature of the 1992 rainfall pattern was the early start of the rains followed by a dry spell from early June until mid July. Following the break, rainfall was above average in August and early September. Rainfall measurements at the SSS and CWSS are shown in Figure 8.30 and Figure 8.31 respectively. The rainfall total at the SSS was 787 mm, whereas at the central east site it was only 496 mm. Over the whole study area, rainfall was 537 mm, only marginally below the 1950-1990 average of 550 mm. No rain fell in the measurement square after 16 September. Consequently, there was the marked delay in the start of the growing season in contrast with an average year. A dry period in June caused the millet crop in most fields sown in May to fail except in the south west. Fields elsewhere were resown, sometimes more than once. In the southern site millet was harvested between 10 and 15 September but not until about 10 October in the central sites.

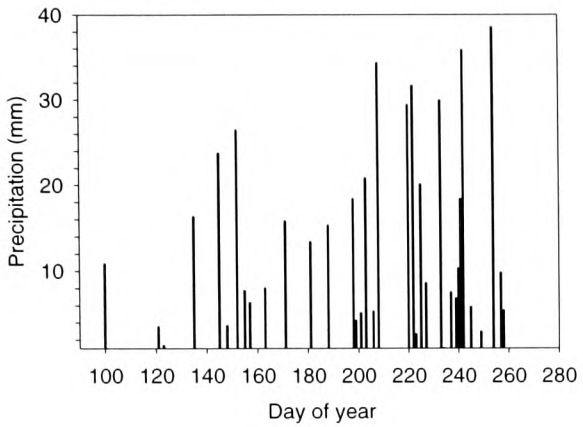


Figure 8.30 Precipitation at the SSS during 1992.

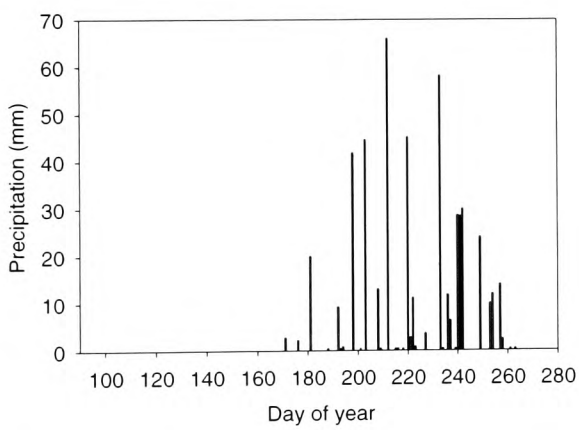


Figure 8.31 Precipitation at the CWSS during 1992.

Remote sensing measurements

NDVI measurements were available from two NOAA AVHRR satellites throughout the growing season. The overall time period of data acquisition was once a day from May to October. There were generally two overpasses per day per satellite at approximately 0200, 0900, 1400, and 2100 GMT. However, not all of these data are processed and distributed because of cloud cover and other data problems. NDVI data were also available on a limited number of dates for both SSS and CWSS from low altitude aircraft measurements. Figure 8.32 shows a comparison of aircraft-derived and satellite-derived NDVI for both sites. Unfortunately, a considerable degree of noise is evident in the satellite NDVI data caused by the relatively coarse spatial scale of the measurements and the heterogeneous

nature of the landscape. Comparison of aircraft and satellite measurements suggest that the upper boundary of the scatter of points delimits the true value, whilst the lower values may be distorted by view angle effects. This poses a problem for model optimization purposes. Therefore, a trained genetic algorithm was used to fit a curve to the raw NDVI data using a profile based on the measured LAI curves for the two sites. The results are shown as the dotted lines in the figures. It is difficult to ascertain the validity of these curves since detailed ground based measurements were not available; however, the shapes of the curves appear to be reasonably realistic.

Unfortunately, for the duration of this study it proved impossible to obtain processed ASAS high-resolution directional reflectance data for any of the HAPEX sites. However, a limited number of miscellaneous ground-based data were generously provided by A. Bégué (see Table 8.7) and these, together with the NDVI data, provide the basis for the remote sensing input in the optimization process.

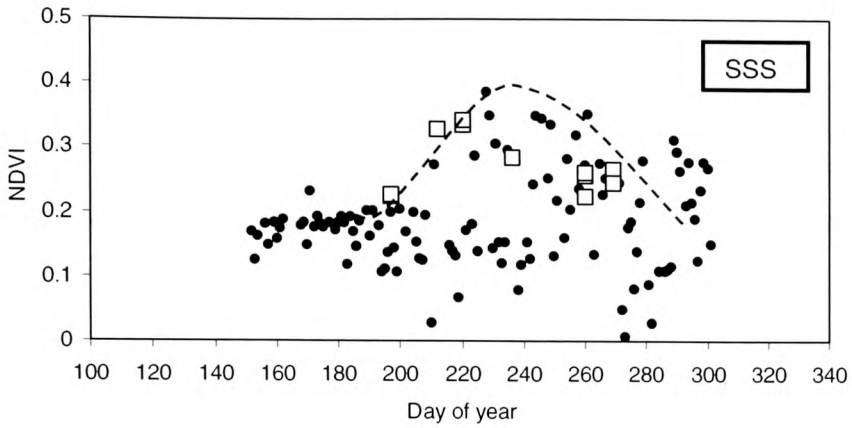
Day of year	*SSS*	*CWSS*
184	✓	
193	✓	
211	✓	
229		✓
230	✓	
250	✓	
242		✓
259		✓
270		✓
273		✓

Table 8.7 The availability of ground based directional reflectance data for millet at the southern (SSS) and central western (CWSS) supersites. (Data supplied by A. Bégué)

8.6.4. Efficiency Strategy Results

The simulated annealing technique was used to optimize a simple light use efficiency model of millet biomass production over the 1992 growing season, as described in the Section 8.6.1 above. A combination of specific heat analysis and trial and error (see Chapter 6) were used to obtain the optimum set of annealing schedule parameters for the optimization: these are shown in Table 8.8.

(a)



(b)

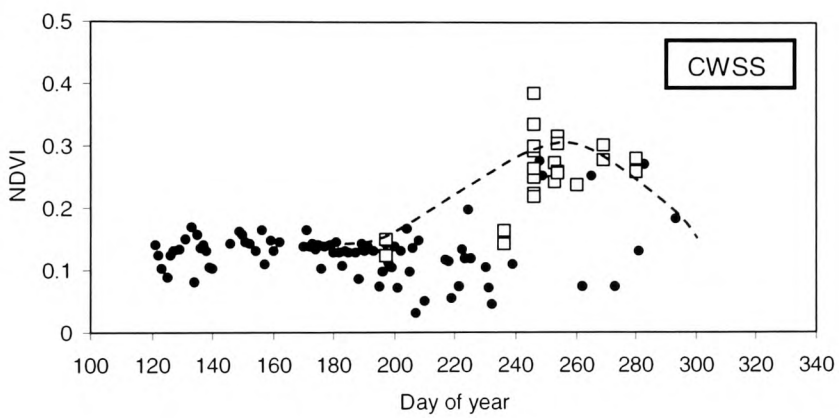


Figure 8.32 Measured NDVI values for, (a) the southern supersite, and (b) the central western supersite. Circles represent AVHRR measurements, squares represent values derived from a Piper Saratoga aircraft data, 200 m altitude. Dashed lines show a curve fitted to the data using a trained genetic algorithm approach where the typical curve shape was derived from measured LAI values at the two sites.

Parameter	*Value*
Initial temperature, T_0	5000
Maximum iterations, K	100,000
Evaluations per level, f_{lev}	100
Attenuation factor, α	1.0
Above ground LUE, ε_{ag}	1.87 g MJ^{-1}

Table 8.8 Simulated annealing parameters used for the optimization of the simple LUE model, and the average for ε_{ag} obtained from the inversion.

Table 8.8 also shows the average value of ε_{ag} obtained from the optimization: ε_{ag} was however not found to be constant over time as illustrated in Figure 8.34. Unfortunately, no measured values for ε were reported in the HAPEX-Sahel dataset, therefore the points in Figure 8.34 represent values of ε calculated from field measurements of APAR and biomass production. Under optimum conditions, the biomass conversion efficiency of C_4 plants is considerably higher than for C_3 species, with typical mean values of around 2.5 g MJ^{-1} for C_4 and 1.93 g MJ^{-1} for C_3 crops. This difference is explained by the reduced photorespiration that is characteristic of C_4 plants, because CO_2 is concentrated at the rubisco reaction sites in the bundle sheath cells.

Figure 8.33 shows the values of LAI and canopy height derived from the remote sensing data during the optimization process, compared with measured values. One can see that the estimations of LAI from remotely sensed data are quite good. The largest errors occur shortly after emergence when the very small leaf area and low canopy coverage make the signal to noise ratio for the remotely sensed data very small. The fit of the canopy height estimates is rather poorer, with again, the largest error occurring shortly after emergence.

Day length and row spacing do not affect ε_{ag} in millet (Monteith *et al* 1983; Azam-Ali *et al* 1984) and whilst temperature may have an impact on the duration of the growing cycle, it does not affect conversion efficiency (Squire *et al* 1984). Consequently, the two main factors responsible for the reduction of ε_{ag} from its optimum value are reduced stomatal conductance resulting from high atmospheric saturation water vapour pressure deficit, and water stress (Ong and Squire, 1984). The millet crops at the CWSS was not irrigated and was known to be experiencing water stress several times during the season because of lack of rain. This is consistent with the decline in the value of ε toward the end of the season as shown in Figure 8.34.

Figure 8.34 shows that the optimization produced a value for ε_{ag} that tended to be higher than the values derived from PPFD interception data measured during HAPEX-Sahel. The average of these measured values was ε_{ag} = 1.6, which is considerably lower than the potential value, for optimium growth conditions. The value obtained from the optimization, ε_{ag} = 1.87 g MJ^{-1} compares favourably with measurements made during the IRI/ORSTROM experiment in Niamey in 1989 (Varlet-Grancher, *et al* 1990) where measured ε_{ag} values for irrigated and unirrigated millet were 2.6 and 1.82 g MJ^{-1} respectively.

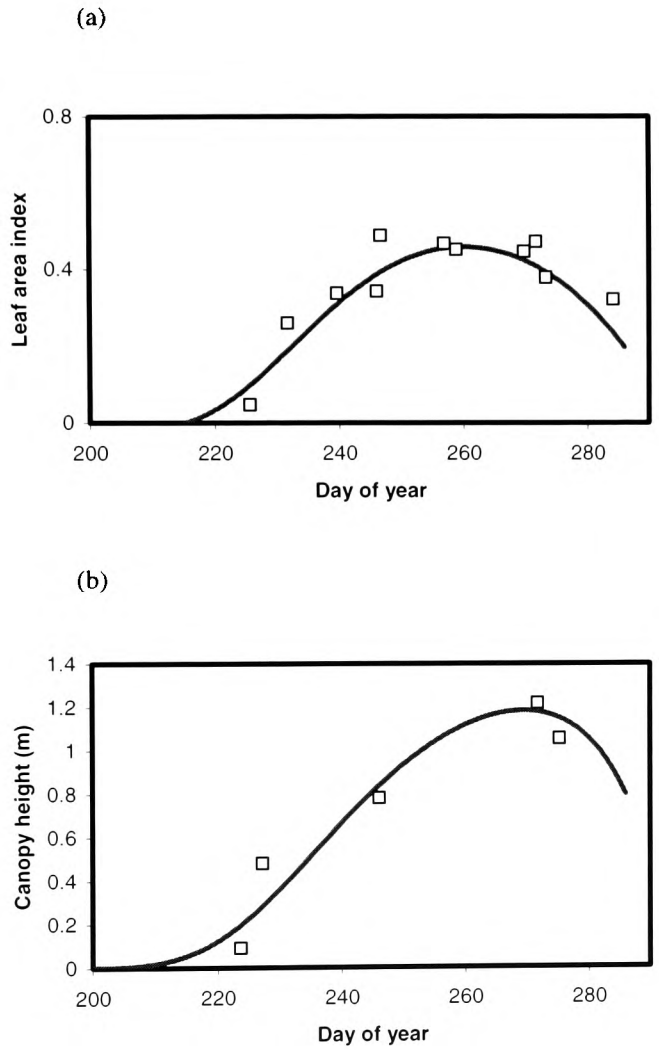


Figure 8.33 The temporal evolution of millet canopy structural parameters at the CWSS during the 1992 growing season: lines – derived from field measurements obtained from the HAPEX-Sahel database, symbols – derived from remotely sensed directional reflectance and NDVI data. (a) LAI and (b) canopy height.

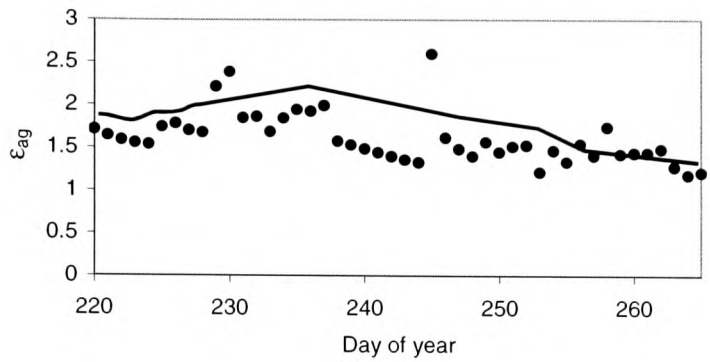


Figure 8.34 Above ground light use efficiency (ε_{ag}) for millet at the CWSS during 1992, modelled using the optimized FRAMEWORK model (line) and calculated from field measurements obtained from the HAPEX-Sahel database. ε_{ag} was calculated from field measurements as follows: daily above ground biomass was obtained from a curve fitted by the investigators (S. Prince *et al.*, University of Maryland) to field measurements of biomass, APAR was measured by A. Bégué. ε_{ag} was then calculated from Equation 8.23.

8.6.5. Allocation Strategy Results

Again, the simulated annealing technique was used to optimize the FRAMEWORK model as described in Section 8.6.2 above, to attempt to fit model predictions of allocation to the measured data. A value of 1.87 g MJ^{-1} was assumed for ε_{ag} as found in the Section 8.6.4. Straight-line relationships were assumed between total leaf biomass and LAI, clump diameter and canopy height between zero and the maximum value for each parameter. The results of the inversion are shown as the dashed grey line in Figure 8.36, Figure 8.37 and Figure 8.38. The fit to the measured data is very good in all cases. As can been seen from Figure 8.36 the predicted allocation of biomass into the leaves (dashed grey line) was very close to the measured value during the early stages of development, but as the maximum LAI was approached there was a tendancy for underestimation.

Figure 8.37 shows that allocation to leaf biomass based on clump diameter is underestimated towards maturity. The relationship between clump diameter and leaf biomass depends largely on canopy architecture and, in particular, on the leaf size, shape and inclination angle. The nearly linear nature of this relationship indicates that as a clump grows its height, leaf area and clump diameter increase but the leaf area density remains nearly constant.

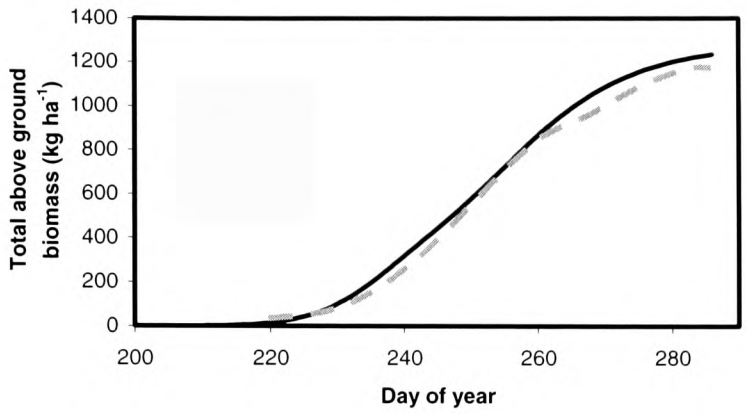


Figure 8.35 Comparison between measured above ground biomass (solid black line) and modelled above ground biomass using the efficiency strategy (dashed grey line) at the CWSS for the 1992 growing season. The measured curve is a fit made by the investigators (S. Prince *et al.*, University of Maryland) to their field measurements.

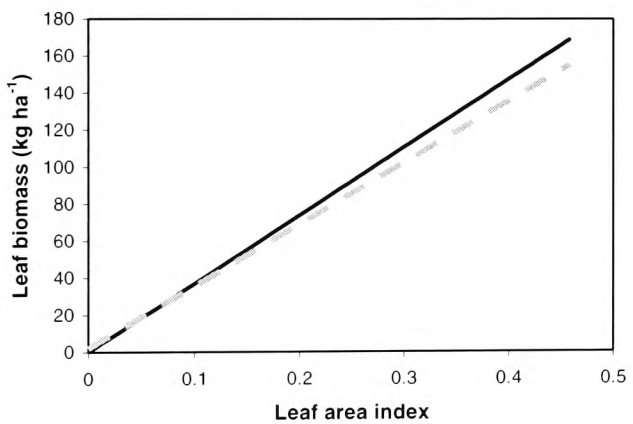


Figure 8.36 The relationship between LAI and leaf biomass for measured for millet (solid black line) during the 1992 growing season at the CWSS. The measured relationship is derived from curve fits made by the investigators to their field measurements. The dashed grey line shows the results of simulated annealing optimization of biomass allocation within FRAMEWORK using an LUE approach with ε_{ag} = 1.87 g MJ^{-1}, assuming a planting density of 3255 clumps ha^{-1}.

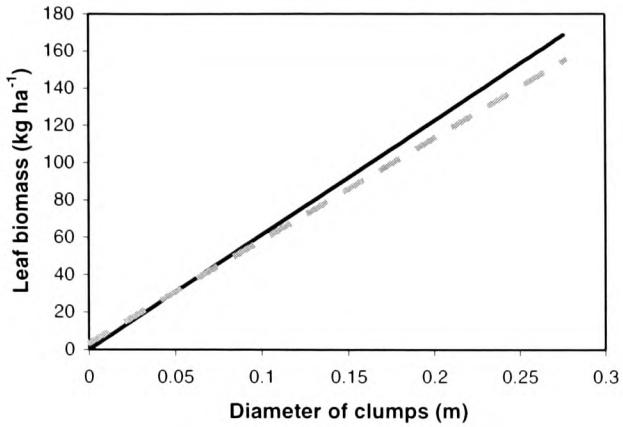


Figure 8.37 The relationship between clump diameter and leaf biomass measured for millet (solid black line) during the 1992 growing season at the CWSS. The measured relationship is derived from curve fits made by the investigators to their field measurements. The dashed grey line shows the results of simulated annealing optimization of biomass allocation within FRAMEWORK using an LUE approach with ε_{ag} = 1.87 g MJ^{-1}, assuming a planting density of 3255 clumps ha^{-1}.

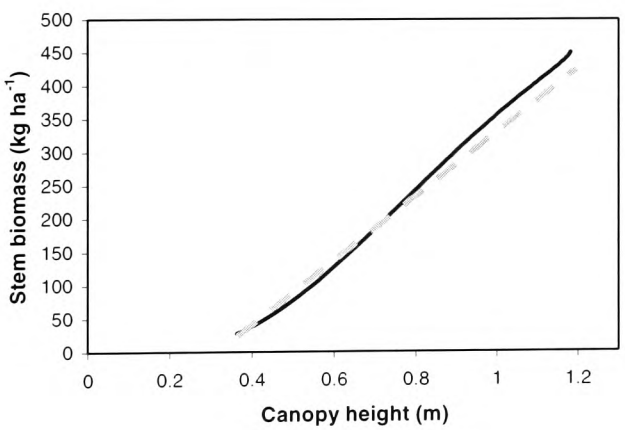


Figure 8.38 The relationship between canopy height and stem biomass measured for millet (solid black line) during the 1992 growing season at the CWSS. The measured relationship is derived from curve fits made by the investigators to their field measurements. The dashed grey line shows the results of simulated annealing optimization of biomass allocation within FRAMEWORK using an LUE approach with ε_{ag} = 1.87 g MJ^{-1} assuming a planting density of 3255 clumps ha^{-1}.

Figure 8.38 shows that the allocation to stem biomass based on canopy height tended to be overestimated in the early stages of growth, but underestimated toward maturity. The measured

relationship between canopy height and stem biomass is slightly less linear than the relation between leaf biomass and LAI. This probably reflects changes in stem structure that are required to support the plants as they grow.

The underestimation of allocation to both leaf and stem biomass suggests that either the value for ε_{ag} was too small, or that systematic errors occurred during the optimization. Estimations of canopy height from remote sensing were only possible using directional reflectance data, not NDVI data. This limited the number of measured data available, which is, in turn, reflected in poorer optimization results. Consequently, it is anticipated that, in general, the estimations of stem biomass will be less accurate than estimations of leaf biomass.

Figure 8.39 shows the incident PAR and the PAR absorption of the optimized model for the growing season in 1992. The PAR absorption shows relatively little day to day variation with only days 245 and 277, corresponding to dust storms, showing a significant departure from the trend. This is reflected in the uniform growth predicted in Figure 8.40.

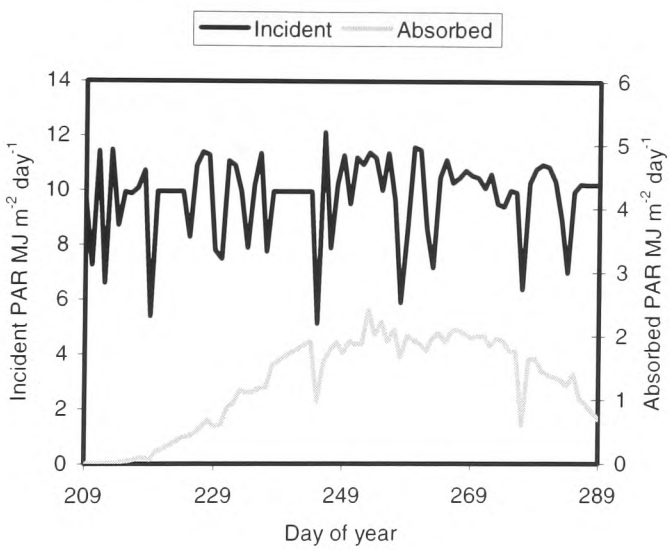


Figure 8.39 Comparison of daily incident PAR and PAR absorbed by millet at the CWSS.

Figure 8.40 shows the temporal evolution of the various biomass components predicted as a result of the biomass allocation optimization. The spike biomass was assumed to be given by:

$$W_{spike} = W_{ag} - W_{leaf} - W_{stem} \tag{8.26}$$

Where: W_{spike} is the spike biomass

W_{ag} is the total above ground biomass

W_{leaf} is the total leaf biomass

W_{stem} is the stem biomass

The results show a tendency systematically to underestimate above ground biomass production in general. This may result from inadequate or inaccurate representation of the surface reflectance by the combined Bégué, PROSPECT and SOILSPECT+ models, errors in directional reflectance or NDVI values, an assumed value for ε that was too small, or a bias in the optimization itself. A bias in the optimization is extremely difficult to quantify but is likely to contribute to the underestimation of biomass production, at least to some degree, because of the inherent complexity of the optimization process. Unfortunately, estimates of errors for the directional reflectance measurements were not available, so it is difficult to comment on their possible influence on the results of the optimization. Figure 8.18 and Figure 8.19 show that the Bégué model represents the visible waveband directional reflectance of millet quite well. Although the representation of near-infrared reflectance by the model is not as good as for visible wavebands (Figure 8.20), it is still reasonable, and the ability of the model to predict NDVI values, illustrated in Figure 8.21, is good. However, the heterogeneous nature of the landscape makes utilization of remote sensing measurements difficult, and explains, to some extent, the considerable scatter in the NDVI values shown in Figure 8.32. It is likely that these data introduce a considerable amount of noise into the optimization process, making parameter retrieval more difficult.

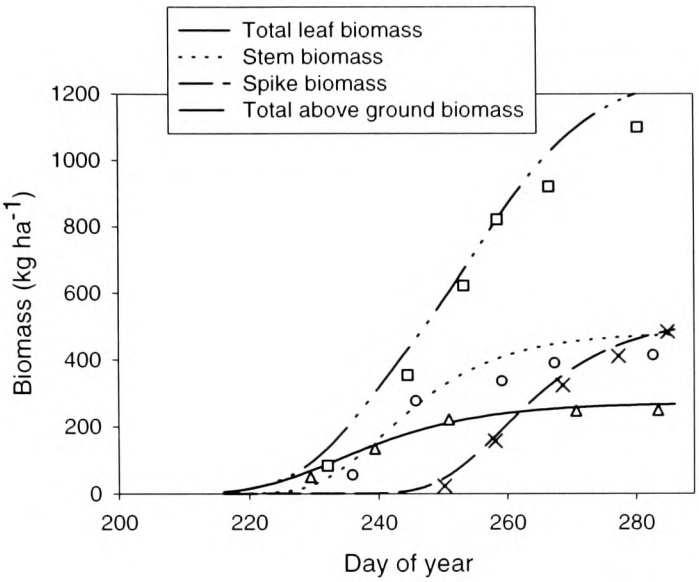


Figure 8.40 Temporal evolution of various millet biomass components at the CWSS, measured (lines) and derived from simulated annealing optimization of biomass allocation within the FRAMEWORK model (symbols): total above ground biomass (□), total leaf biomass (△) stem biomass (○), spike biomass (×)

Although there is a systematic underestimation of both biomass production and allocation, the magnitude of the underestimation is quite small (Table 8.9). Although the error for spike biomass production is only 3%, this may be rather fortuitous since the underestimation of production in the other components is in proportion and spike biomass is derived from Equation 8.26 above.

Component	*Measured* (kg ha^{-1})	*Estimated* (kg ha^{-1})	*Percentage error*
Total above ground biomass	1235	1118	-11%
Total leaf biomass	268	243	-8%
Stem biomass	476	395	-17%
Spike biomass	490	480	-3%

Table 8.9 A comparison of measured component biomass with predictions from the allocation optimization, for millet at the CWSS.

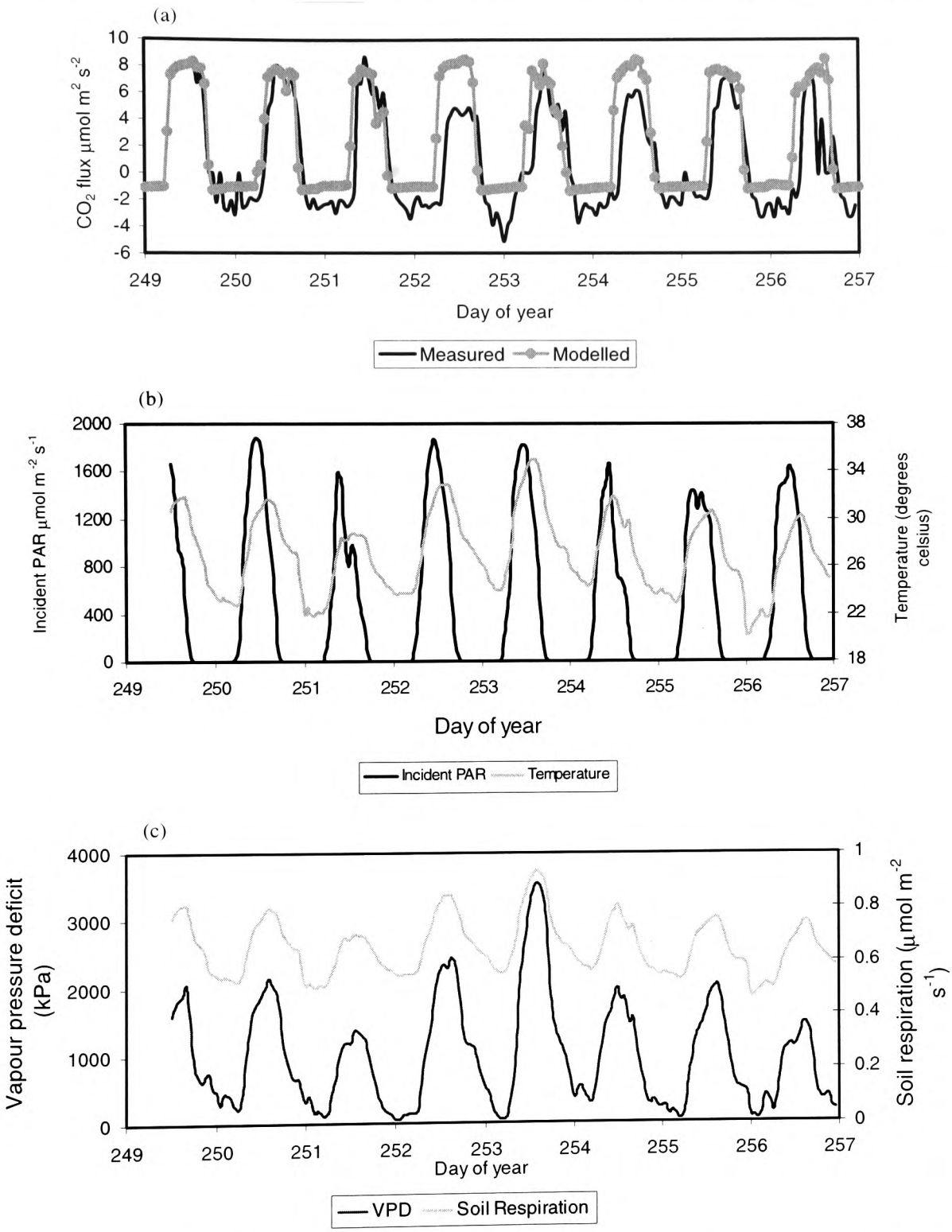


Figure 8.41 (a) Net ecosystem CO_2 flux for millet at the SSS modelled using the CRCB model within FRAMEWORK, and measured using eddy covariance – positive fluxes indicated carbon assimilation. (b) Incident photosynthetically active radiation (PAR). (c) Atmospheric saturation vapour pressure deficit calculated using the FRAMEWORK model, and soil respiration flux calculated as described in Section 8.5.4, (see Table 8.4, and Figure 8.26).

Coupling the LUE and Mechanistic Models

To investigate the coupling of the LUE and mechanistic models within FRAMEWORK, the allocation strategy was applied to the millet crop at SSS. The biomass allocation was then fed into a mechanistic model based on the CRCB model as described in Section 8.5.5. Hourly CO_2 assimilation was predicted and allocated on a daily basis according to the optimized allocation functions, by assuming that the biomass contained 47% carbon by weight (Penning de Vries, *et al* 1989).

Nine days of eddy covariance data were available for validating the CRCB model within FRAMEWORK. The data were measured by a team from the University of Edinburgh comprising J.B. Moncrieff, S.L. Scott, J.M. Massheder and P.G. Jarvis (Moncrieff *et al.* 1997). The closed-path system was mounted on a mast 9 m above the ground with a fetch of 1500 m in all directions with approximately 90% of the signal emanating from within 400 m of the mast (J.B. Moncrieff *pers. comm*). The measurement period ran from 5 to 12 of September (Julian day 249 to 256) at which time senescence had started and the leaf area index was 0.46. EddySol software was used to derive CO_2 flux measurements from the raw data with appropriate correction for density fluctuation, sensor separation and signal attenuation being applied (Moncrieff *et al.* 1997).

The CRCB model was parameterized, using leaf scale CO_2 flux data measured by P. Levy, University of Edinburgh, as described in Table 8.5. Figure 8.41a shows a quite close agreement between the measured and modelled fluxes. The diurnal trend in flux is extremely well simulated, with a similar predicted pattern for each day, reflecting the similar course of radiation temperature and humidity for each day. The sharp increase in VPD between days 251 and 253 was not reflected in the CO_2 measurements and the small measured flux on days 252 and 254 cannot be accounted for in terms of the measured environmental variables. The eddy covariance system was still in prototype form when the measurements were made and was subject to a number of minor technical problems that may account for this discrepancy.

The model-predicted nocturnal respiration fluxes were almost constant. Measured values were significantly more variable, and the model underestimated night-time respiration by a little over 1 $\mu mol\ m^{-2}\ s^{-1}$ on average. There is considerable uncertainty in the soil respiration predicted by the model since it is based on measured leaf respiration parameters and the fraction of biomass in the roots. It is assumed that the respiration resulting from microbial activity in the soil is negligible (see Section 8.5.4): this seems a reasonably assumption because of the very small organic matter content of the soil. The soil respiration predicted by the model is rather small (Figure 8.41c), with an average of 0.75 $\mu mol\ m^{-2}\ s^{-1}$. This is approximately half the soil respiration measured at the fallow site shown in

Figure 8.25. This discrepancy many explain a large amount of the underestimation of night-time fluxes, however there may be other contributory factors: the model was parameterized using leaf respiration data, but other plant components may contribute significantly to night-time respiration. Variability in nocturnal fluxes of forests has been reported by several authors (e.g. Grace *et al* 1995, Hollinger *et al* 1994). This has been attributed to storage within the canopy air space. Nocturnal conditions are typically characterized by low wind speeds and atmospheric stability. Under these conditions there is little turbulent mixing and CO_2 concentrations may build up in the air mass below the sensor. This CO_2 is sporadically flushed by large, infrequent eddies. Unfortunately, corrections for storage effects do not always remove the nocturnal variability in measured fluxes, and a full explanation of this phenomenon is, as yet, unknown.

The CRCB model was run for the entire growing season on an hourly time step in order to predict the carbon assimilation and allocation into above and below ground components. Figure 8.42 shows the incident and absorbed PAR and daily maximum and minimum temperatures for millet at the southern supersite. Figure 8.43 shows the predicted daily and cumulative net fluxes during the growing season. Carbon assimilation follows the general trend of PAR absorption and leaf area development shown in Figure 8.42 and Figure 8.44. The large CO_2 flux predicted by the model is typical of a fast growing C_4 crop.

Two sets of millet biomass production measurements were available for comparison with FRAMEWORK predictions (Table 8.10). Model predictions are comfortably within the confidence limits, differing from measurements by -5.4% on 15 August and +5.2% on 15 September. The progression from underestimation of biomass in August to overestimation in September may be attributed to several factors:

- a shift in biomass allocation between above and below ground components as the plants reach senescence, possibly in response to changes in soil water or nutrient availability,
- inaccuracies in the estimation of canopy reflectance and APAR as the crop reaches maturity,
- changes in the balance between photosynthesis and respiration that are known to occur in the latter stages of plant development, and
- the loss of plant parts during senescence which were not included in the 15 September measurements.

Component	Measured (kg ha^{-1})		Modelled (kg ha^{-1})	
	15 Aug	15 Sep	15 Aug	15 Sep
Above ground biomass production	2580	2730	2499	2800
Below ground biomass production	500	650	416	755
Total biomass production	3080 ± 650	3380 ± 190	2915	3555
Grain yield		640		

Table 8.10 Comparison of measured millet biomass production and production modelled using FRAMEWORK at the SSS during 1992. Modelled production assumes the biomass consists of 47% carbon by weight. Data obtained from HAPEX-Sahel database. Data for 15 August measured by P. Levy, University of Edinburgh; data for 15 September measured by J. Brouwer, ICRISAT, Niger.

Figure 8.45 shows comparisons between modelled and measured soil water content and soil water potential at the SSS. The modelled results for only two soil depths are shown for clarity. There is a large amount of scatter in the measured data reflecting the heterogeneous nature of the site. The figure shows that the simple four-layer bucket model used is capable of representing the changes in soil water content quite well, and identifies the key features in a satisfactory manner. For example, the surface (20 cm) soil layer is much more sensitive to transient rain events than the deeper (60 cm) layer. The pattern of surface layer soil moisture closely follows the rainfall pattern shown in Figure 8.30, while the pattern in the 60 cm layer indicates a degree of inertia, and has an attenuated response. These features are clearly present in the model.

It should be noted that measurements of bulk density were not made at this site. Bulk density can have a considerable effect on the calibration equation. An average bulk density was assumed although one might expect the bulk density to increase considerably in the laterite, which starts 0.2 to 0.5 m below the soil surface. Therefore, although the data are presented here as water contents, as values these should be treated sceptically. However, bulk density has only a small effect on the gradient of the calibration equation, so calculation of changes in water content with time should be reliable.

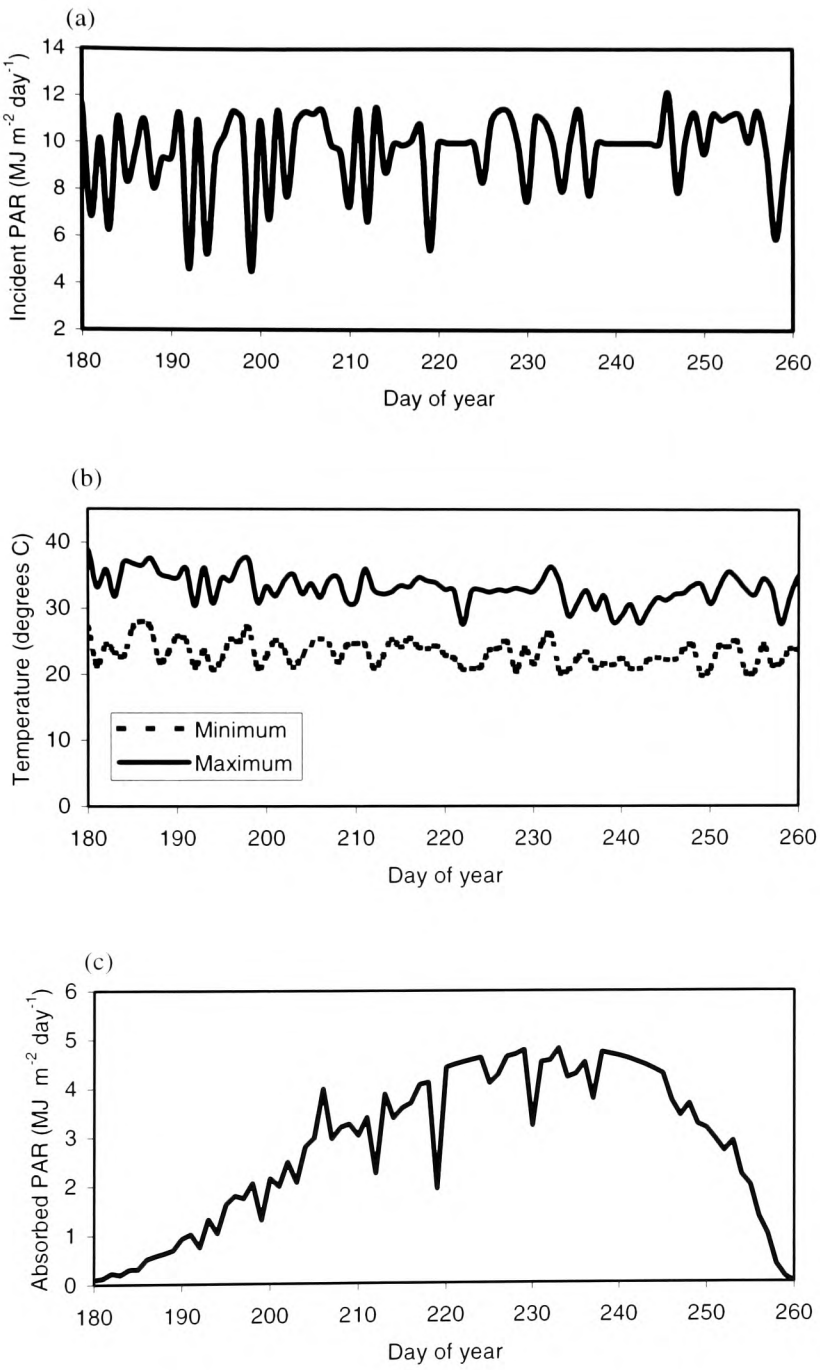


Figure 8.42 (a) Daily incident PAR in MJ per day, (b) daily maximum and minimum temperatures , (c) PAR absorption modelled by the FRAMEWORK, for millet at the SSS during the growing season in 1992. Data for graphs (a) and (c) were measured by A. Bégué, S. Prince and N. Hanan, University of Maryland. Data for graph (b) were measured by B. Monteny *et al.* ORSTOM, Laboratoire d'Hydrologie, Montpellier. All data obtained from the HAPEX-Sahel database.

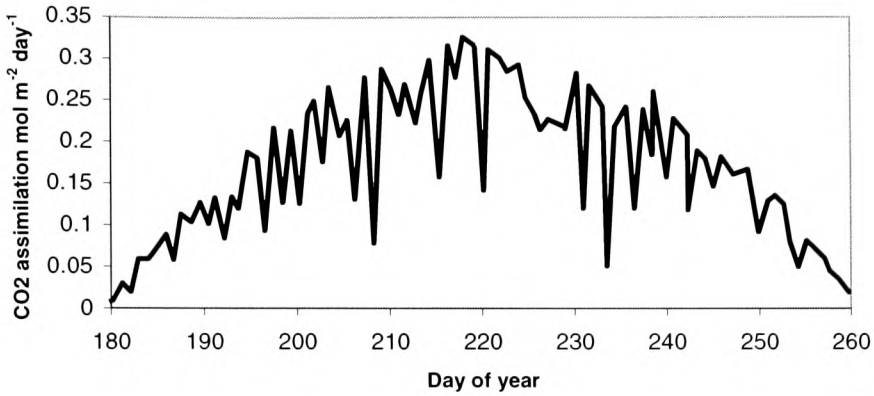


Figure 8.43 Daily net CO_2 flux for millet, modelled using FRAMEWORK.

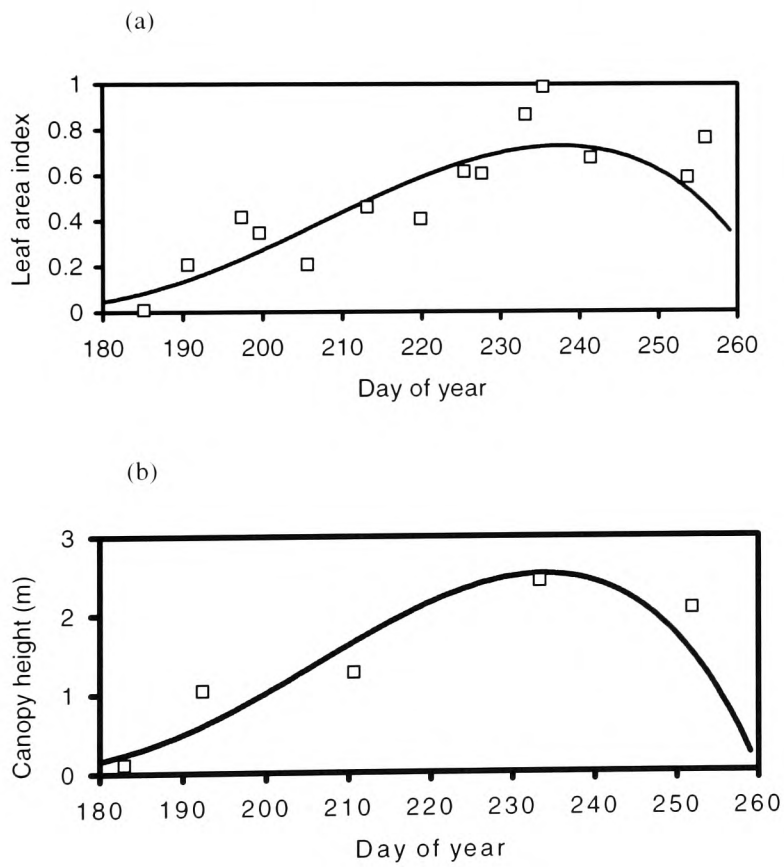
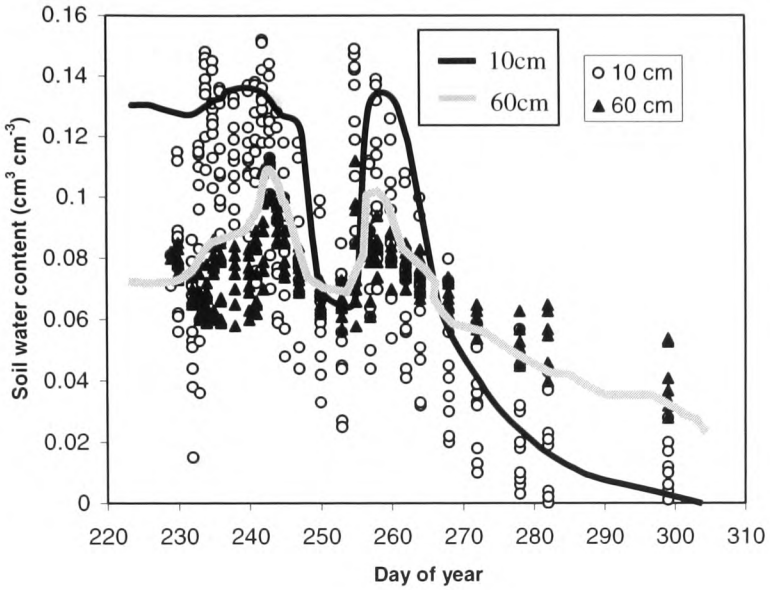


Figure 8.44 The temporal evolution of millet canopy structural parameters at the SSS during the 1992 growing season: (a) LAI and (b) canopy height. Lines field measurements, symbols estimated from remotely sensed data. Data were measured by S. Prince, University of Maryland. (Data obtained from the HAPEX-Sahel database)

(a)



(b)

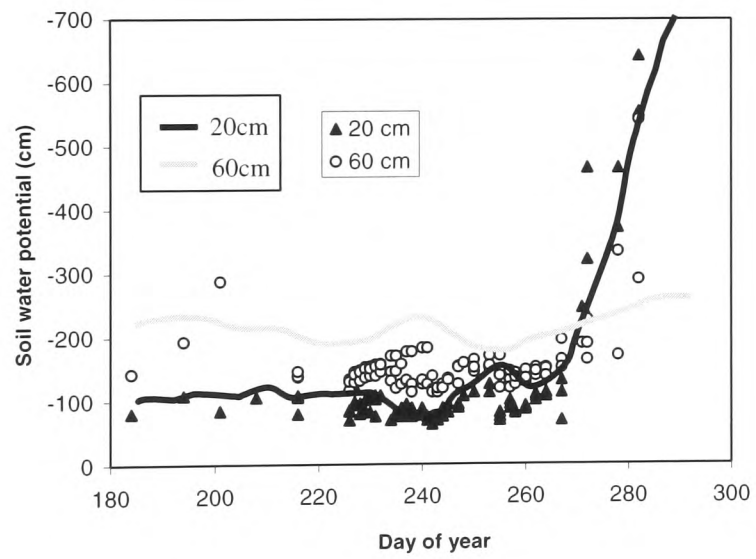


Figure 8.45 Comparison of, (a) soil water content and (b) soil water potential measured at four different depths at the SSS, with predictions from the FRAMEWORK model. Measurements made by J.D. Cooper, Institute of Hydrology.

8.7. CONCLUSIONS

Object-oriented programming techniques provide a powerful tool for developing highly flexible, configurable models. The FRAMEWORK model has been successfully configured to model millet, by linking the CRCB C_4 photosynthesis model and a Jarvis-type stomatal conductance model with the Bégué model of radiative transfer in regularly clumped canopies.

The Bégué model was capable of simulating the radiative transfer in a millet crop to a reasonable degree of accuracy in both visible and infrared wavebands, showing that, even in the near infrared, the approximation of single scattering is valid for sparse canopies with small leaf areas and a bright soil background. The CRCB model, when configured within FRAMEWORK, was able to predict both leaf scale and canopy scale CO_2 fluxes with reasonable accuracy. This indicates that the theory on which the model is based is sound, as it apparently copes well with the range of leaf area and environmental variables encountered over the season.

Optimizing hourly time step models over a period of more than a few days is impractical because of the computational effort required. Two strategies for optimizing FRAMEWORK using the validation-optimization approach and remotely sensed data were successfully implemented. In the efficiency strategy, a simple light-use-efficiency model was optimized using remotely sensed data. The value of ε_{ag} estimated from this approach was toward the lower end of values reported in the literature, indicating that the crop was suffering environmental stress. Unfortunately, no field measurements of ε were available, but rainfall patterns suggested that the crops was under water stress for a least part of the growing season. This was reflected in the small leaf area and canopy height, and in the small yields following the harvest.

The allocation strategy optimized light use efficiency to estimate the allocation of biomass into above ground components, assuming that ε_{ag} was known. The estimated component biomass values derived using the allocation strategy were, in general, slightly underestimated, however, the magnitude of the errors was small and comfortably within typical measurement error. Several factors may have contributed to the errors, for example the chosen value for ε_{ag} may have been too small. Systematic errors are common during complex optimizations, moreover, only a limited number of directional reflectance measurements were available, hindering the optimization process.

Light use efficiency and mechanistic models were successfully coupled within FRAMEWORK by employing the allocation strategy to estimate above ground biomass allocation based on optimization of a LUE model.

The CRCB model was configured within FRAMEWORK using the estimated allocation functions and used to model above and below ground biomass production for the growing season. The agreement between modelled and measured biomass production shown in Table 8.10 suggests that the FRAMEWORK model works well, at least where the parameters can be measured accurately. The accuracy of the FRAMEWORK biomass estimations enables confidence in the model's flux predictions, and this is further reinforced by the comparison with eddy covariance data shown in Figure 8.41. Because the carbon content of the vegetation was not measured there is an inherent uncertainty in the comparison between modelled and measured biomass estimates. This problem would not arise if modelled fluxes could be compared with measurements made over the whole of the growing season. Unfortunately, such data were not available because of problems with the prototype flux measuring system.

To make long-term predictions using mechanistic SVATs would require growth to be simulated dynamically, rather than prescribed with constant parameters. This, in turn, would necessitate the use of mechanistic models of assimilate allocation that are a function of plant nutrient status, and so would require simulation of nutrient cycling. The techniques described in this Chapter may provide a useful way of validating such models, although a more thorough validation of the techniques would be required. Long-term eddy covariance measurements would provide better validation data for the model and a more complete soil respiration model is also desirable. The functions used here are based on a fit to measured data and depend solely upon temperature; respiration resulting from microbial activity was assumed to be negligible. Whilst this assumption may be valid for the HAPEX-Sahel site where soil organic carbon content is very small, this assumption is unlikely to hold for soils with a larger amount of soil organic carbon.

Chapter 9

Improving Model Performance

In Chapter 5 the contrast between mechanistic and empirical vegetation models was described and in Chapter 8 models were further classified into big-leaf and multi-layer models. Mechanistic multi-layer models are complex and may involve a heavy computational burden. Traditionally, for stand-scale modelling, it has been necessary to strike a compromise between the level of mechanistic detail within the model and the computing time required to perform calculations by making simplifying assumptions and approximations with regard to vegetation function. Conventional serial computers are steadily increasing in power and are thus able to perform more calculations ever more quickly. However, parallel computing techniques are becoming more widely available to researchers and may offer the potential for a large increase in performance provided vegetation models can be written to take advantage of parallel programming techniques. This chapter introduces parallel computing, and discusses the problems involved with designing parallel SVATs.

9.1. PARALLEL COMPUTING

Serial computing on Von Neumann style machines is characterised by a single processor that executes a series of instructions in order to produce a result. Parallel computing is concerned with producing the same results using multiple processors. Therefore problem to be solved must divided up amongst the processors. There are a variety of ways in which this can be done and this process is critical to the performance of the parallel program.

9.1.1. Parallel Architecture

Two main types of parallel processor architecture exist and are classified according to Flynn's Taxonomy, which labels hardware by instruction stream and data stream. A conventional serial computer is classified as Single Instruction Single Data (SISD) as a single processor executes one instruction on one piece of data in a single clock-cycle. Figure 9.1 shows the classification of the main types of parallel architecture according to Flynn's Taxonomy.

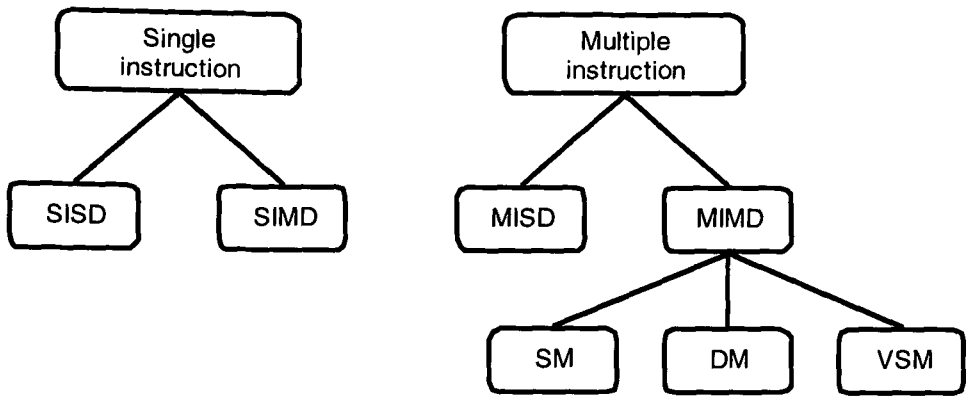


Figure 9.1 The classification of parallel processor architectures according to Flynn's Taxonomy.
SISD – single instruction single data
SIMD – single instruction multiple data
MISD – multiple instruction single data
MIMD – multiple instruction multiple data
SM, DM, VSM – shared, distributed and virtual shared memory configurations.

Single instruction multiple data (SIMD)

SIMD architecture typically consists of many simple processors, each with local memory in which it keeps the data to work on. Each processor simultaneous performs the same instructions on its local data, with instructions issued by the controller processor. In addition processor can communicate with each other, in order to perform shifts and other array operations. SMID architectures are good for problems where the same operation is performed on a number of different objects; a good example of this is image processing. However, if the loads on the processors are not balances poor performance can results because the execution is synchronised at each step.

Multiple instruction multiple data (MIMD)

MIMD architecture consists of several independent processors capable of executing individual instruction streams (i.e. individual programs) simultaneously. Three sub-types within the overall definition of MIMD exist:

Shared memory

In this architecture, there are usually a small number of processors, each with access to a global memory store via some sort of bus (Figure 9.2). Inter-processor communication is handled by one processor writing information to a location in the global memory, and another processor reading the data. This has the major advantage that it is easy to program, as there are no explicit communications between processors. Access to global memory is handled using techniques for traditional multi-tasking computers, e.g. through the use of semaphores. Unfortunately, the shared memory

architecture does not scale well. When the number of processors becomes large, a bottleneck develops when several processors attempt to access the global memory at the same time.

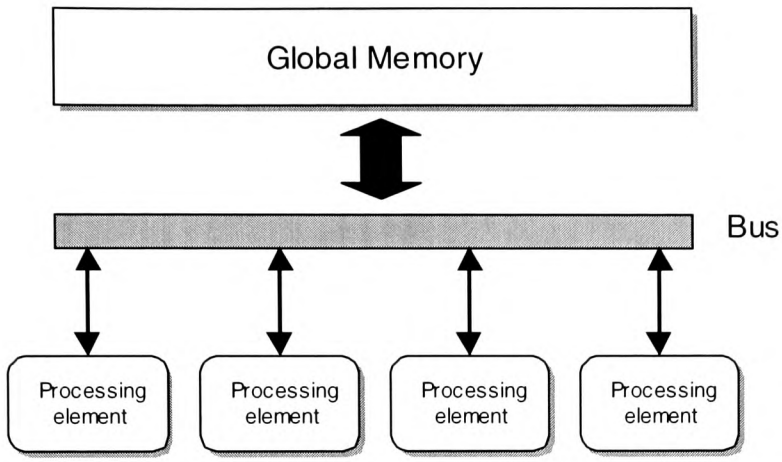


Figure 9.2 Shared memory architecture

Distributed memory

In the distributed architecture the shared memory bottleneck is avoided by giving each processor its own memory Figure 9.3. A processor can only access that memory that is directly attached to it. If a processor requires information that is held in the memory of another processor, it must send a message to that processor requesting the data. Clearly, access to the local memory is very much faster than access to data on a remote processor, which will be longer the greater the physical distance involved. This non-uniform access time is affected by the way the processors are inter-connected. As the number of processor increases, it quickly becomes impossible to connect a processor to every other processor. There are several possible solutions to this problem. One method is to connect together a processor to a subset of its neighbours. This allows messages to be sent from one process to another via a number of intermediate processors. Inter-connecting processors are often laid out in a hypercube arrangement because this has the advantage of not increasing radically the number of connections as the number of processors is increased. Current design practice is to connect processors to network routing chips rather than directly to other processors themselves. The same topology rules apply, but the processors no longer play any part in message forwarding, this allows much greater design freedom.

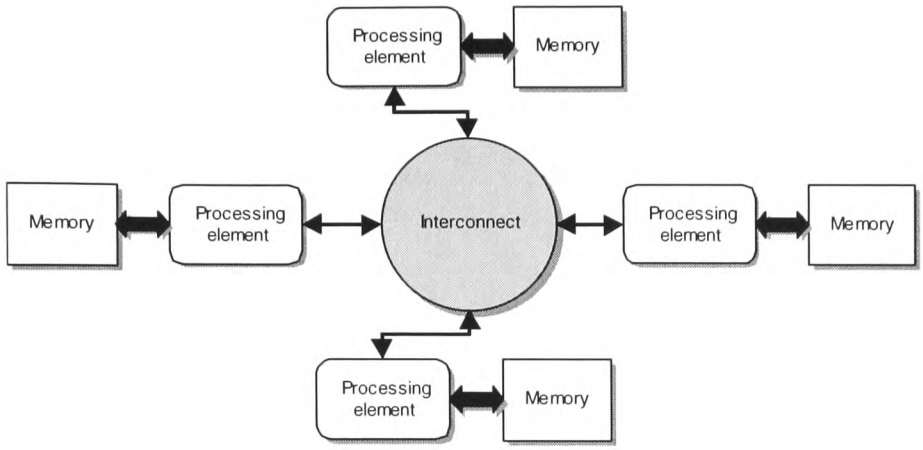


Figure 9.3 Distributed memory architecture

Another, extremely valuable, type of distributed memory machine is in fact clusters of workstations. In such "machines" the processors of the individual workstations are connected via Ethernet networking. While clusters are more loosely coupled than custom made single machines they provide a relative cheap way of developing parallel code which can later be ported to custom parallel machines, thus saving development time on expensive hardware. The development of HPC languages such as MPI and FORTRAN 90 greater facilitate the creation of parallel code and the use of workstation clusters for parallel processing, because they avoid the need to rewrite code when porting to different machines.

Virtual shared memory

The above classifications are in a sense idealised architectures. Actual machines usually conform to a combination of architectures, an example being Virtual Shared Memory machines such as the Cray T3D. In these machines each processor has its own memory like a distributed memory machine, but direct access can also be made to a remote global address space. This is possible because of the incorporation of dedicated circuitry to deal with communications independently of the processors.

9.1.2. Decomposing

Consider a sequential program that is to be ported to parallel hardware (in particular distributed memory machines) in order to gain a performance increase. Sequential programs are generally composed of inherently sequential parts, and potentially parallel parts. Somehow the work of the potentially parallel parts must be split up, or decomposed, in such a way that a number of processors

can work on the problem in parallel. However, decomposition itself frequently introduces an overhead (for example message passing between processors) so that any gain by parallelizing the code must outweigh any overheads. Three main decomposition techniques exist, and these are shown in Figure 9.4

Trivial decomposition

This term is used to describe problems that do not really require decomposition at all. Consider a sequential program that is to be run independently on lots of different sets of input data. Clearly, by performing each run on a different processor simultaneously, an almost linear increase in speed can be achieved if each run takes approximately the same time. Since each run completes independently, no communication is required and therefore this technique can be used very effectively on workstation clusters

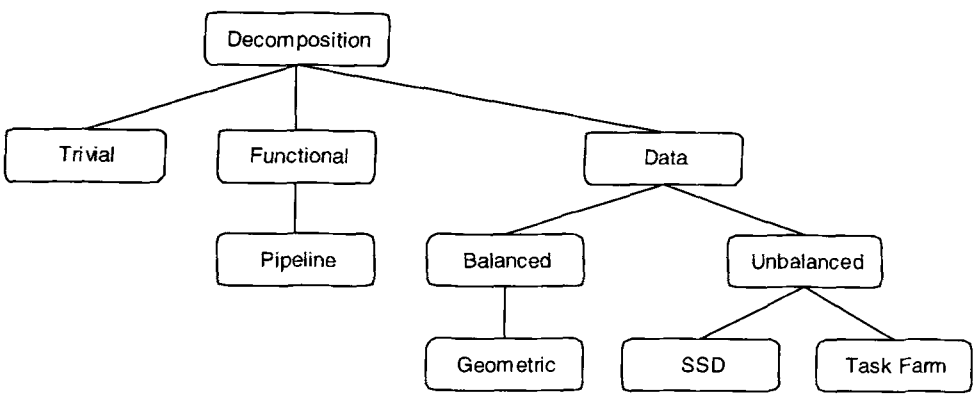


Figure 9.4 Decomposition techniques.

Functional decomposition

Function decomposition breaks up the program into a number of sub-programs. A simple example of this is the pipeline Figure 9.5 in which each input passes through a number of subprograms in turn. Parallelism is introduced by moving several inputs through the pipeline simultaneously. This technique is widely used in modern microprocessor design and is partly responsible for the increase in CPU performance in recent years. Parallelism in the pipeline is limited by the number of stages, and for maximum efficiency, all of the stages need to be kept busy. This requires that the time taken for each stage is the same - the pipeline is then balanced.

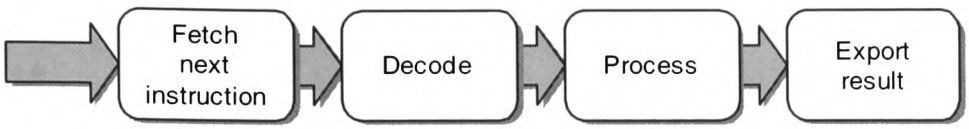


Figure 9.5 An example of a four-stage pipeline from the fetch-decode-execute cycle of a microprocessor.

In general the effectiveness of this technique is problem dependent and the degree of parallelism is limited by the program. This means that the individual elements in a large dataset are unlikely to be processed any faster than those in a small dataset, consequently a large dataset takes a proportionately longer time to process.

Data decomposition

Many problems involve applying similar operations to different parts of a large dataset. In such cases, rather than decomposing the program itself, it may be possible to decompose the data. A "processing grain" is used to refer to the work involved in processing a number of basic data items. The grain size is the number of basic items in the processing grain. In most cases, in order to compute the result at one point in the dataset corresponding data from other points (normally neighbouring points) are required. If the processing grains consist of clusters of data points the results for these points can be calculated without reference to the rest of the dataset in exactly the same way as a sequential program. However, the processing of boundary points within the grain requires the gathering of information which typically resides on other processors. The gathering of this information wastes valuable processing time proportional to the distances over which the information has to travel. Thus, a conflict exists between the desire to maximize the number of processors to exploit parallelism, and the communications overheads involved with splitting up the dataset. The most important of these overheads are the communications costs, leading to the requirement to maximize the ratio of the time spent on useful calculation, to the time spent on communications. This is the so called "calc-to-comms" ratio.

Geometric data decomposition

Since the number of processors is usually fixed, the calc-to-comms ratios is maximized by dividing the dataset into grains and allocating one grain to each processor. As discussed above, not every

processor is connected to every other so it is advantageous to allocate neighbouring grains to neighbouring processors, thus reducing latency. This is known as geometric decomposition.

Simple geometric decomposition works well provided each grain takes the same time to process. However, in some applications, this is not the case and a situation may arise where some processors have no work to do at all! A general way to solve this problem is to use small grains, and allocate more than one grain to each processor. There are two strategies for this, scattered spatial decomposition (SSD) and task farming. In scattered spatial decomposition a random selection of grains is allocated to each processor, trusting that, on average, each processor will have an approximately equal workload. There is a direct trade off between the relatively larger amounts of communication associated by having smaller grains, and the balance between the work loads of different processors. This is known as the "load balancing" problem (Figure 9.6). The scattered spatial decomposition is a "static" load-balancing technique because once the grains have been allocated, no further attempts are made to even-out the workload between processors. Figure 9.6 shows an example of the load balancing problem from image processing: the task is to identify the edges in an image. A simple geometric decomposition with few processors (Figure 9.6a), leaves some processors with a lot of work and other with little to do, thus the problem is unbalanced. The standard solution to this problem, shown in Figure 9.6b, is to use smaller grains.

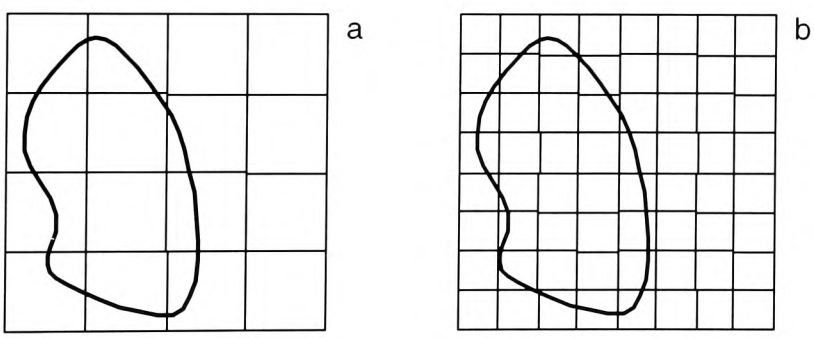


Figure 9.6 Geometric decomposition of an unbalanced problem, a) with large grain size and b) with small grain size.

If the grains can be processed totally independently of each other then task farming can be used to provide dynamic load balancing. One processor is identified as the task master. The other worker processors repeatedly request grains from the task master, process it and dispatch grains to a collector, which may or may not be part of the task master. This has the major advantage that no prior assumptions need to be made about the dataset: provided that the grains are sufficiently small, an even workload is ensured, since a processor that happens to receive many grains requiring little work, merely requests more grains. However, the costs associated with maintaining the dynamic load balance can become significant when many iterations are being performed over the dataset. In such a situation, static allocation requires only a single transmission of data from the master to the workers, while a task farm involves re-transmission for every iteration.

Message passing concepts

Parallel processing achieves performance gains by utilising co-operation between processors. This means that parallel program design involves both a definition of the processes that will be executed by the processors, and a specification of how individual processes are to synchronise and exchange data with one another. In message passing languages such as MPI, processors communicate by exchanging messages. Normally, a single source code program is written, compiled and linked on a front-end computer. The resulting object code is then copied onto each processor. Thus, every processor interprets the same object code. Parallelism is achieved by placing processor specific branches in the source code, so that certain logical paths are only executed on certain processors.

9.2. "PARALLELIZING" THE FRAMEWORK MODEL

Unfortunately decomposing vegetation models is a difficult task since hour-on-hour (or day-on-day, or whatever time step is chosen) simulation is an inherently sequential process. At any given time vegetation function is largely a function of the ecosystem state in the previous and earlier time steps. In addition to the question of temporal scale, spatial scale considerations also have a bearing on model decomposition. Consider for example modelling vegetation at the landscape scale, where different vegetation functional types are represented in a patchwork structure. For a given time step, models could be run within each patch in parallel. Although inter-process communication would be required for each time step (for example, to integrate fluxes from each patch into the boundary layer) the processing time required to run each model within its patch would maintain a reasonable calc-to-comms ratio. This type of model utilises a trivial, event-base decomposition model but requires inter-process communication after each iteration, and this lends itself to a combination of a geometrical data decomposition approach with task farming (Figure 9.7). If the number of patches exceeds the available number of processors, then each processor must calculate the results for more than one

patch. Patches may contain widely different vegetation types, which may require different vegetation models in order to represent their function accurately. This has serious implications for the distribution of tasks by the master processes because one processor (or group of processors) must be confined to working on one type of patch. This, in turn, has implications for load balancing, with maximum performance occurring when there are equal numbers of different types of patches. If a processor were to work on type types of patch that require different models one after the other then either the models would have to be combined into one huge program, or the program would have to be dynamically loaded on to the processor. Neither of these is desirable in performance terms and is likely to negate any performance gains achieved through parallel processing.

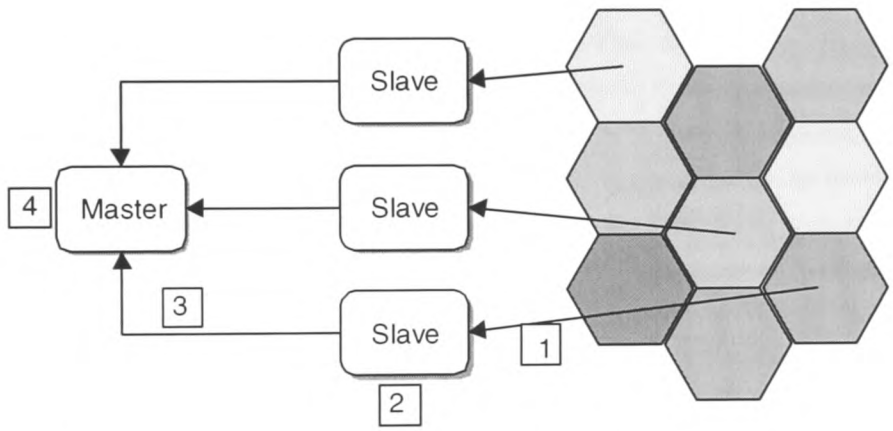


Figure 9.7 Decomposing a patchwork vegetation model to run on a parallel computer. Different patch shading indicates different vegetation types.
(1) The master process assigns a patch to each slave processor.
(2) The slave runs the vegetation model for one time step.
(3) The slave returns the result to the master, which then assigns the slave a new task.
(4) The master process collates the results from each patch.

The FRAMEWORK model overcomes many of these problems by using a single flexible model structure that is able to represent the function of many different vegetation types. The flexibility of FRAMEWORK is achieved largely through the use of object-oriented programming techniques and the C++ language. Unfortunately, a message passing implementation of C++ has not yet been developed, and consequently, in order to develop a message passing version of the model suitable for running on a variety of parallel hardware, it was necessary to a re-write the model in MPI-C. This is quite a time consuming task and in order to avoid spending too much time on this task, just the model kernel, and a number of selected methods were ported to MPI-C. The scope of this Chapter is not to

derive detailed results from the parallel FRAMEWORK model (P-FRAME) but to study the potential for the use of parallel computing techniques in vegetation modelling as a whole.

9.2.1. Parallel Computing Hardware and Software

One of the benefits of MPI over earlier parallel processing languages is its portability. A version of the MPI compiler can be installed on most workstations, in addition to dedicated parallel machines such as the Cray T3D. Workstations provide an extremely cost effective way of designing parallel code both in terms of the development time and the run time on expensive parallel platforms. MPI code can be developed on a workstation, using *simulated* processors, i.e. although the workstation has a single processor the compiler is able to simulate any number of other processors in software. This gives no performance benefit (in fact the maintenance overhead of the software is likely to reduce performance relative to sequential code), but it does enable code to be debugged without using up expensive and valuable CPU time on machines such as the Cray T3D. Moreover, machine such as the Cray T3D are essentially batch processors, where jobs are scheduled to run at specific times without user interaction. This greatly extends the compile-debug-run cycle of code development, since errors in the code are only found following a batch run. Once a bug-fix has been applied, the developer must await the results of the next batch run, which may not take place for a number of days, to discover if the fix was successful. In contrast, code can be developed interactively on workstations and compilation repeated until all errors are removed.

The Institute of Ecology and Resource Management has two Sun workstations connected to the departmental Ethernet: 1 SPARCstation IPX with 32MB RAM and 1 SPARCstation "Classic" with 24 MB RAM. CHIMP MPI was installed on these machines and the P-FRAME code was developed, first on a single machine using simulated processors, then using the two workstations as a distributed parallel processing machine. The developed code was then ported to a number of different parallel computing machines for benchmarking. I am very grateful to a number of groups for generously supplying computing time: University of Cambridge, Kilean Systems and ICL. Table 9.1 lists the computing hardware used for benchmarking P-FRAME in various configurations. Available computing time on each machine was severely limited, consequently it was not possible to run call configurations of the model on all machines. Benchmarking was therefore restricted to calculating the relative speed-up of parallel P-FRAME running on multiple processors compared with a single processor. No inter-machine performance comparisons were made.

	Machine	*Configuration*
A	2-processor Sun cluster	1 Sun SPARCstation IPX; 1 Sun SPARCstation Classic workstations
B	8-processor Sun Cluster	8 Sun SPARCstation 5 workstations connected to a 100Mbit s^{-1} LAN
C	Cray T3D	512 DEC Alpha processors, used in various domain sizes of up to 8
D	Cray CS-6400	16 SPARC processors used in various domain sizes of up to 8

Table 9.1 Parallel computing hardware used for benchmarking P-FRAME

9.3. IMPLEMENTING P-FRAME FOR A MULTI-PATCH LANDSCAPE

P-FRAME was configured to model the carbon assimilation for a patchwork of vegetation like that shown in Figure 9.7. An arbitrary landscape of seven patches was used to evaluate the performance of the model. These patches consisted of three different vegetation types: a prairie grassland using data obtained from the FIFE experiment (Sellers *et al* 1992), a Sitka spruce stand using data collected by J. Massheder, University of Edinburgh, and a wheat field using data measured by Wiegand *et al* U.S. Department of Agriculture, Weslaco Texas. The choice of patches was not intended to represent a "real" landscape, and no validation of model predictions was attempted because the intention was purely to estimate the "speed-up" gained by using parallel hardware. Although this was a relatively simple exercise, a measurement of the speed-up obtained by using parallel processing techniques provides a valuable insight into the feasibility of this modelling approach.

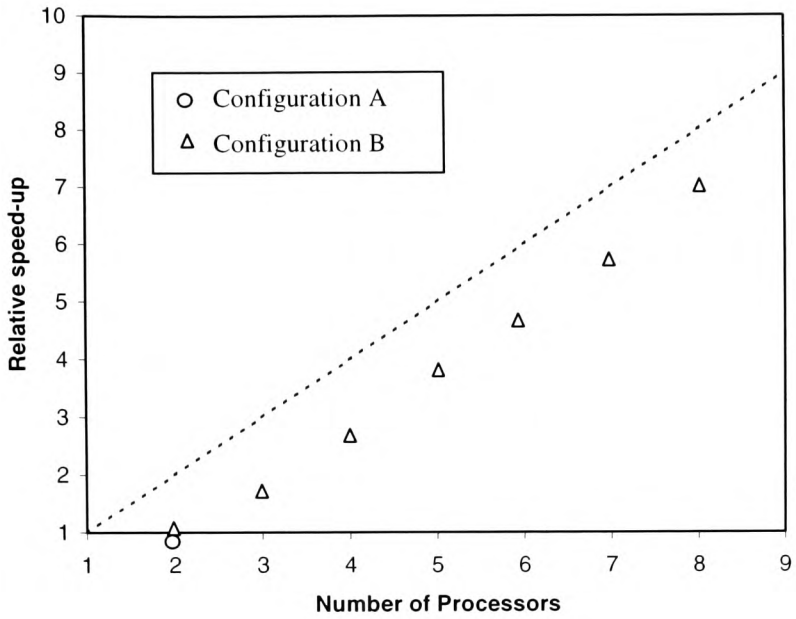


Figure 9.8 Relative speed-up for P-FRAME configured for a seven patch landscape. Configurations A and B correspond to Table 9.1.

Figure 9.8 shows the relative speed-up obtained when running P-FRAME for a seven-patch landscape on two of the parallel hardware configurations listed in Table 9.1. The speed-up obtained with configuration A was only 26%: several factors may contribute to this quite small performance improvement. Firstly, the two machines were relatively far apart, and connected to different loops on the LAN. Secondly, benchmarking was performed during the day when there was a significant amount of network traffic: the performance at times of lower network load (e.g. at night) may have been slightly better, but it is unlikely that any improvement would be more than a few percent. Thirdly, the speed-up was measured relative to the SPARCstation IPX machine. The second machine, a SPARCstation Classic, although having a theoretically superior specification (excluding its smaller memory), was found to be significantly slower that the IPX system. This will result in a slight bias toward a smaller speed-up.

The above problems do not apply to configuration B, where all machines were of the same type and connected to the same spur of a high speed network. As can be seen from Figure 9.8, the speed-up obtained with configuration B, while less than linear, was approximately proportional to the number of processors. This sub-linear speed-up was not unexpected, bearing in mind the communications overhead incurred when individual processors communicate via the LAN. In addition, there is an overhead incurred with maintaining the master-slave relationships, because the master processor may

not be fully occupied at all times. This is illustrated by the slightly better relative performance with eight processors, becuase with this configuration, one processor can be assigned to each of the seven patches, and the eighth used as the master processor.

9.4. MULTI-COMPONENT PARALLELISM

Although the vegetation models running within a patch are essentially sequential in operation, there may still be scope for parallelization depending on how the model is configured. There are two areas of the FRAMEWORK model where it seems likely that inherent parallelism could be exploited: the component structure of the model and the **radiation/BRDF** object.

The aim of using parallel computing techniques in vegetation modelling is to improve model performance by enabling complexity and refinement to be added to models by reducing the execution time to practical levels.

A multi-component configuration provides an obvious opportunity for parallelism because the model effectively loops over each component, calculating the stomatal conductance, assimilation rate and fluxes of CO_2 and water vapour for each component in isolation, before the contributions from each component are combined (Figure 9.9). However, a problem exists because the radiation regime for each component must be calculated, which will depend upon the diffuse radiation transmitted or scattered, and the direct radiation attenuated by other components (see for example Appendix G). This will be true for all but the sparsest canopies. One possible solution is to calculate the radiation regime for every component (in serial) prior to any other calculations carried out in parallel, but this carries the significant disadvantage that no parallel speed-up is obtained for the frequently complex radiation calculations. Alternatively the radiation regime for each component may be calculated in parallel by calculating the direct radiation penetration and diffuse transmittance to the position of the given component within the canopy. The complexity of this calculation will depend upon the specific radiation transfer equations and the inherent approximations that are employed. In general the second solution provides superior performance and was used in this study. The radiation interception or BRDF objects themselves provide another opportunity to exploit parallel processing because models of radiation interception and reflectance are frequently based around a series of loops, iterating over illumination and observation zenith and azimuth angle.

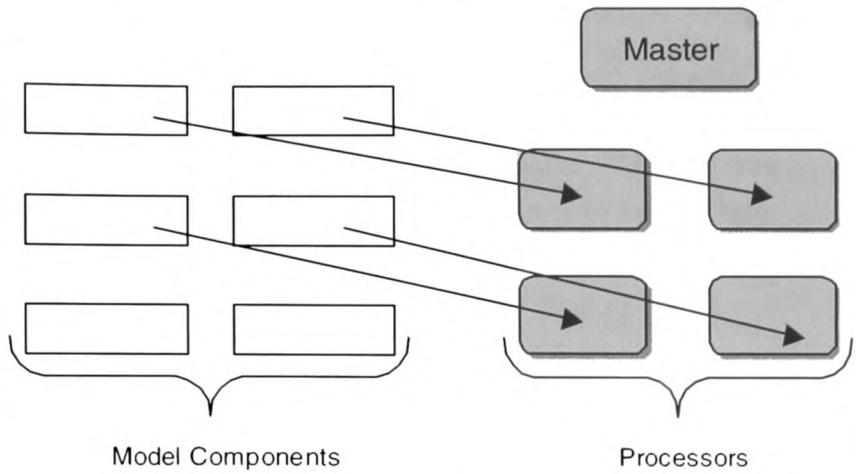


Figure 9.9 Parallel processing of a multi-component vegetation model. The model is run for each component in parallel on separate processors. The master processor collects and collates data from each slave processor. The master also assigns work to each slave.

The performance gain achieved by running the individual components of a multi-component SVAT in parallel was evaluated using P-FRAME. The model was configured to predict the carbon assimilation and canopy fluxes for a Black Spruce (*Picea mariana*) forest in Canada using data collected during the BOREAS experiment. The speed-up for a seven component, multi-layer canopy representation, with a varying number of processors, was estimated using hardware configuration D shown in Table 9.1.

In order to model the light regime with the canopy, and consequently within each model component, new light interception routines based on the analysis by Norman (1978) were incorporated into P-FRAME. Routines based on these equations are also used by the MAESTRO model (Wang and Jarvis 1991). In MAESTRO, these routines calculate the radiation interception and absorption within the crown of a "target" tree of a stand. Included in this calculation is the probability of interception by other crowns within the stand. In this study, the canopy was assumed to be composed of a uniform stand of trees with regularly spaced stems occupying a grid. The leaf angle distribution is assumed to be constant throughout the crown, however, leaf area density varied both vertically and horizontally. Direct and diffuse radiation fluxes are calculated in visible, near infrared and thermal wavebands following the formulation of Norman (1978). For direct radiation, the path length in the direction of the radiation is calculated, weighted by the projected leaf area in that direction. This path length is calculated in two steps, within and outside the "target" tree crown. The latter step involves calculating the entry and exit positions of the light beam through the other crowns in the stand, using the method of Norman and Welles (1979). A mathematical description of these routines in given in Appendix G.

Figure 9.10 shows that the radiation interception model works very well when parameterized using measured data obtained from the BOREAS Information System (BORIS). There is a slight underestimation of APAR around noon on the 4 and 5 September. These days correspond to a cloudy period with small overall values of incident flux, but a large value for the diffuse radiation fraction. This suggests that the model may slightly underestimate diffuse radiation interception. A similar underestimation of PAR toward noon on 2 September is more difficult to explain, as it did not correspond to a period of significantly higher levels of diffuse flux.

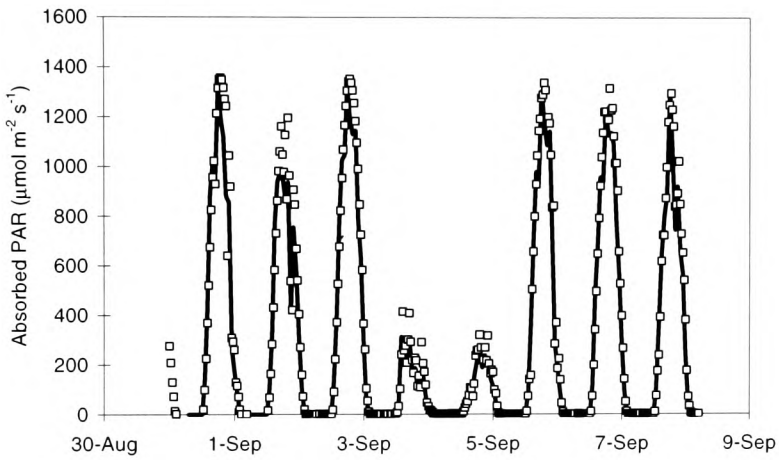


Figure 9.10 Comparison of APAR modelled by P-FRAME for Black Spruce in BOREAS 1994, with measured APAR.

Photosynthesis was modelled using the Farquhar and von Caemmerer (FvC) model (Farquhar, von Caemmerer, and Berry 1980), which is described in Appendix B. A Jarvis-type model was used to predict stomatal conductance using parameters obtained from BORIS. Additional data for parameterizing the photosynthesis and stomatal conductance routines were kindly provided by B. Kruijt, University of Edinburgh. CO_2 fluxes were measured by a team from the Institute of Ecology and Resource Management, University of Edinburgh, lead by Prof. Paul Jarvis. The FvC model was initially parameterized with the default P-FRAME parameter set shown in Appendix F. The canopy nitrogen distribution was then optimized to give the best fit between modelled and measured flux data. This was achieved by varying the parameters shown in Table 9.2 using the genetic algorithm routine within P-FRAME.

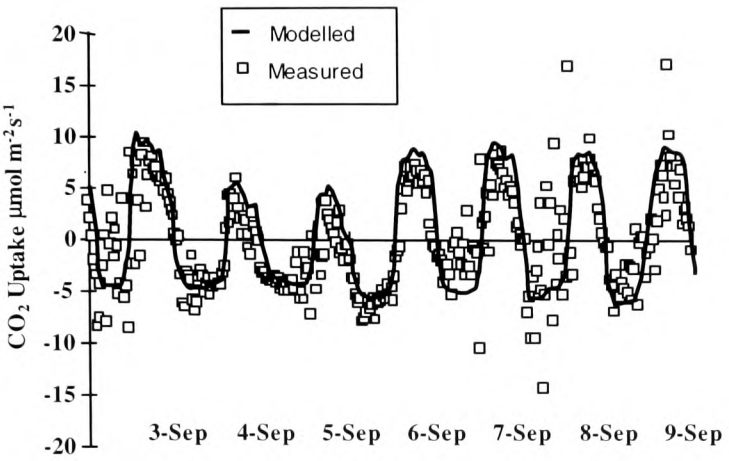


Figure 9.11 Measured CO_2 flux for the BOREAS old black spruce site during seven days in 1994, compared with predictions made by P-FRAME configured with the FvC photosynthesis model.

Parameter	*Value*	*Units*
Slope of J_{max} v N curve	46.4×10^{-6}	mol g^{-1} s^{-1}
Intercept of J_{max} v N curve	-2.4	g m^{-2}
Calibration temperature	27	°C
Slope of V_{cmax} v N curve	28.5×10^{-6}	mol g^{-1} s^{-1}
Intercept of V_{cmax} v N curve	-5.3	g m^{-2}
Calibration temperature	27	°C
Average leaf nitrogen concentration	0.93	g m^{-2} leaf
E_v	53×10^3	J mol^{-1}
E_j	55×10^3	J mol^{-1}

Table 9.2 The relationship of some of the photosynthesis parameters of the FRAMEWORK model with leaf nitrogen content, obtained from a genetic algorithm optimization of the fit of the modelled CO_2 flux to measured flux data.

The optimized, modelled canopy CO_2 flux compares well with the measured values, as shown in Figure 9.11. The diurnal trend in flux is accurately represented although there is a tendency for the day-time CO_2 uptake peak to be too broad. There was also a tendency for night-time fluxes to be under estimated, a phenomenon observed by several researchers (see Chapter 8) for which there is no consensus on interpretation, at present. The decline in canopy flux in the afternoon of 3, 4 and 5 September is represented by the model, but the day-time fluxes on other days tend to be symmetrical around local solar noon. The under estimation of APAR by the model on 4 and 5 September does not appear to be reflected in the canopy scale CO_2 flux estimation.

9.4.1. Multi-Component Speed Up

Figure 9.11 illustrates that the multi-component configuration of P-FRAME represents the CO_2 flux for a black spruce forest well. For complex clumped canopies such as forests, multi-component models allow a more realistic representation of the processes than big-leaf models. However, because of their complexity, multi-component models are both much more difficult to parameterize, and require significantly greater computing resources. Figure 9.12 shows the relative speed-up obtained by running the canopy components of the P-FRAME model in parallel (Figure 9.9), using a differing number of processors.

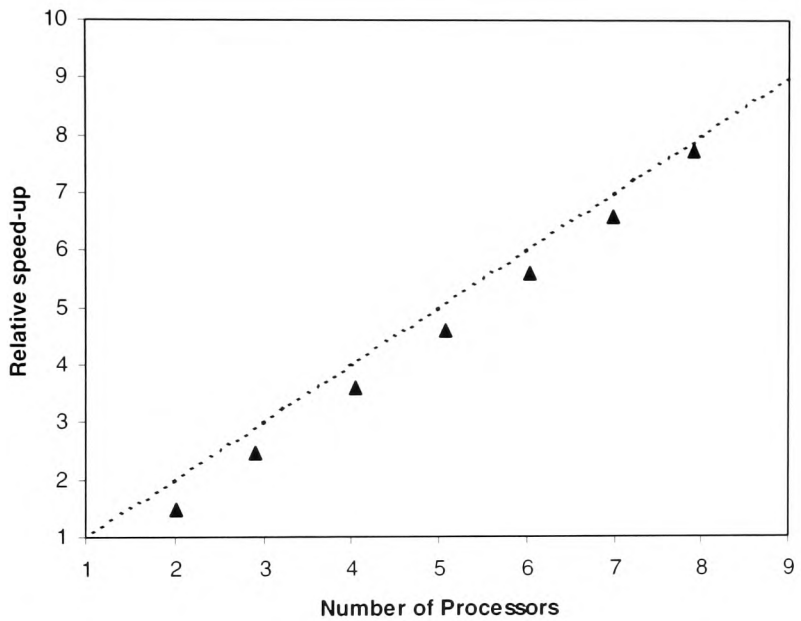


Figure 9.12 Relative speed-up for a seven component configuration of P-FRAME running on a Cray CS-6400 parallel computer, using different sizes of processor domain (i.e. different numbers of processors).

The results in Figure 9.9 are broadly similar to those for the parallel multi-patch configuration shown in Figure 9.8. The relative speed-up is sub-linear, but increases uniformly with the number of processors. The larger relative speed-up obtained with an eight processor domain may be explained by the overhead incurred by using a master processor, Figure 9.9. When eight processors are available each component may be assigned to an individual processor and the eighth processor used as the master. This contrasts with the situation when there are fewer than eight processors where the master must assign more then one component to one of more processors. Overall the speed-up obtained with a given number of processors shown in Figure 9.9 is slightly better than the corresponding result in Figure 9.8. This might be expected beacuse the Cray CS-6400 is a true parallel computer with highly-optimized inter-processor communications and a high speed bus, as

opposed to a network of Sun workstations communicating via TCP/IP protocols on a local area network.

9.5. PARALLEL GENETIC ALGORITHMS

With the FRAMEWORK model there is another area in which parallel processing techniques may be exploited – optimization. Stochastic optimization techniques such as simulated annealing and genetic algorithms are powerful, but computationally very intensive. Genetic algorithms are inspired by natural, biological processes that are inherently parallel. In natural populations millions of individuals exist, reproduce and mutate in parallel. This suggests a degree of parallelism that is directly proportional to the population size used in the genetic search. However, exploiting this inherent parallelism is not necessarily a straightforward task because of the need to maximize the calc-to-comms ratio described in Section 9.1.2. The simple canonical genetic algorithm described in Chapter 6 is not well suited to parallel execution in its basic form. This is because the identification of the fittest individuals and the pooling of results from crossover operations running in parallel involve a great deal of inter-processor communication.

The most direct way to implement a parallel GA is to modify the selection rules of a simple canonical algorithm to introduce tournament selection (Goldberg and Deb 1991). Tournament selection implements a noisy form of ranking. The creation of a new generation involves two steps (Chapter 6): (1) the selection of a number of individuals that are copied to form an intermediate population and (2) the application of crossover and mutation operators to produce the next generation. Instead of using fitness-proportionate reproduction, as used in the roulette wheel method, tournament selection holds competitions (or tournaments) between randomly selected pairs of individuals to determine the intermediate population. This works as follows: two individuals are selected at random from the population, they are compared and the fitter placed in the intermediate population to await crossover. This process is repeated until the intermediate population is full.

With the addition of tournament selection, a parallel form of the canonical genetic algorithm can be implemented in a fairly direct fashion. For example, assume that the population size is N, and that there are $N/2$ processors, then each processor holds two independent tournaments to select two individuals for the intermediate population. Crossover subsequently takes place in parallel.

In the simple canonical GA the population is assumed to be unstructured in the sense that the likelihood of a given pair of individuals mating depends solely on their fitness. However, recently

distributed genetic algorithms (DGAs) have been proposed that impose a spatial structure on populations so that the likelihood of two individuals mating depends on their locality as well as their fitness (Tanese 1987; Whitley 1993). The original motivation for the development of DGA was to exploit parallel computer hardware, and whilst this is one benefit of DGAs, it appears that structured populations also tend to improve the effectiveness of the GA search (Tanese 1989). In unstructured populations it can be difficult for different characteristics to emerge simultaneously, because small differences in the relative fitness of one characteristic over another will cause one to out-compete the other.

Island model GAs exploit a more coarse grain parallel solution than the canonical GA with tournament selection described above. In island models, the population is broken down into sub-populations. Each sub-population is assigned to a processor and a genetic algorithm is executed for each sub-population in isolation. Occasionally (typically every five generations), a few individuals from different sub-populations are swapped – this is known as the *migration* phase. Migration allows sub-populations to share genetic material. Each independent genetic algorithm search will be slightly different because the initial sub-populations will impose a certain sampling bias, moreover genetic drift will tend to drive the sub-populations in slightly different directions (Turner 1987). When migration occurs some characteristics, developed in isolation on one island, will combine with a different characteristic developed on another island. Sampling errors and genetic drift are particularly significant factors for small populations sizes. In island model DGAs, the only inter-processor communications occur during the migration phase, which takes place at a regular interval.

These concepts are not new; Wright (1932) tried to explain how populations in nature are able to escape local optima and discover new, advantageous genes. In general, reproduction and mutations result in properties of sub-populations drifting randomly around some local fitness optimum. If one sub-population were to drift into the influence of a different, higher fitness optimum, then it will be pulled up towards this new peak through natural selection. This sub-population will produce a large number of offspring of higher fitness which will then migrate into other sub-populations, spreading the new genes. Theoretical studies have shown that a relatively small amount of migration is sufficient to disperse advantageous genes through a population (Phillips 1993), but it is unclear whether the necessary combination of migration, selection and drift occur in nature for this process to take place. However, these ideas provide a useful basis for optimization through the medium of parallel processing.

9.5.1. Parallel DGA Speed-Up

The FRAMEWORK genetic algorithm optimization routines were modified to run on parallel computing hardware by implementing the DGA concepts described above. The speed-up obtained by running the P-FRAME DGA on hardware configuration C in Table 9.1 was estimated by optimizing the "bump" function described in Chapter 6. The DGA requires two additional parameters to the canonical GA, the migration interval i, specifying the number of generations between each migration, and the number of individuals that migrate, m (usually expressed as a percentage of the total population).

Parameter	*Value*
Number of generations (n_p)	100000
Population size	250
Survival probability (p_s)	0.8
Crossover probability (p_c)	0.8
Mutation probability (p_m)	0.0004
Number of dimensions (n)	50
Migration interval (i)	10
Migration percentage (m)	20%
Maximum obtained	0.787

Table 9.3 Parameters for the DGA optimization of the bump function and the maximum value of the function obtained by the optimization.

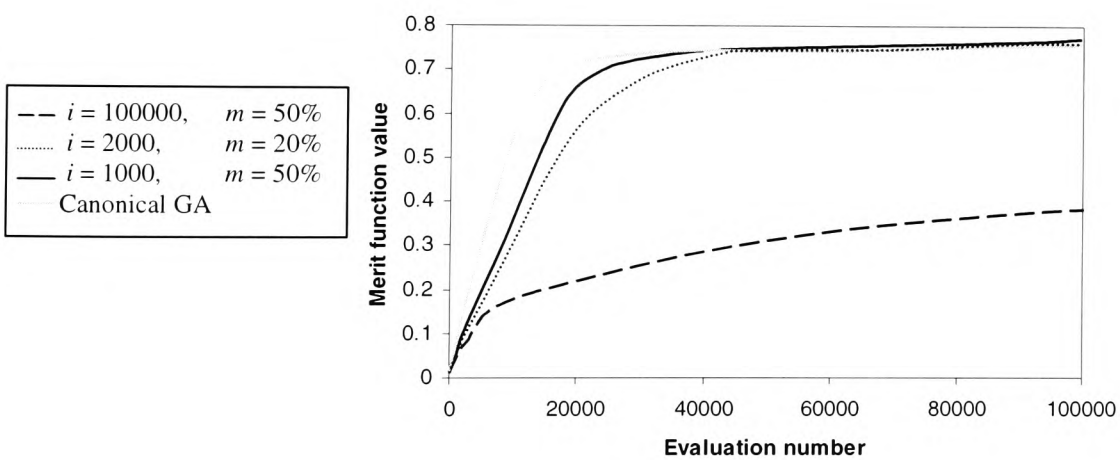


Figure 9.13 Results of the distributed genetic algorithm (DGA) optimizations of the bump function for three different values of the parameters i and m, compared with the performance of the canonical GA.

Table 9.3 shows the DGA parameters used in the optimization, and the resulting maximum of the bump function that was obtained. Note that the maximum function value of 0.787 is better than that

reported in Table 6.7 (0.767) obtained using the canonical GA. This result demonstrates that DGA are powerful optimization tools and appears to confirm that structured populations tend to improve the effectiveness of the GA search. The performance of the DGA algorithm was found to be sensitive to the choice of i and m, as illustrated by Figure 9.13. The DGA with $i = 100{,}000$ is equivalent to a partitioned GA, with populations developing in isolation, merging shortly before termination (Tanese 1987, 1989). In contrast to the hypercube migration scheme of Tanese (1987), the recipient and donor sub-populations were selected at random for each migration. Figure 9.13 shows that the partitioned GA ($i = 100{,}000$) performed worst of all, in contrast to the findings of Tanese (1987) (who optimized a rather different set of function) where the partitioned GA performed best of all. The best performance in this study was achieved with $i = 1000$ and $m = 50\%$. This result is surprising since $m = 50\%$ is a very large amount of migration, whereas Tanese (1987) found $m \approx 10\%$ gave optimum results. In this study, trials with $m = 10\%$ performed less well than other migration rates, and also took longer to find the optimum. The performance of the algorithm declined for values of i much smaller than 1000. This was probably a result of depletion in diversity as the sub-populations converged at higher migration frequencies (i.e. shorter time intervals between migrations).

These results suggest that the optimum values of i and m may be application specific, although further studies with a wide variety of different functions would be required before this statement could be made with confidence. It is known that some trials using the canonical GA are characterised by rapid increases in population fitness, followed by periods of stagnation (Muhlenbein 1991). It may be possible to reduce these periods of stagnation by adopting a DGA in which the migration parameters vary with the number of evaluations made.

From the above it seems probable that with the right choice of migration parameters DGA can at least match the performance of the canonical GA, but what speed-up is achievable with the DGA? Figure 9.14 shows the results of the parallel optimization with $i = 10$, and $m = 20\%$, averaged over five runs. As can be seen from the figure, a linear speed up was obtained using these parameters. In fact, the measured speed-up with five or more processors was super-linear. This was not a feature of the averaging process, since all five runs returned super-linear speed-up values. This result is difficult to explain but it may be that with five or more processors the workload was sufficiently divided that each processor was able to store its entire workload within its own sub-cache, rather than the core cache of the machine, thus speeding up memory access.

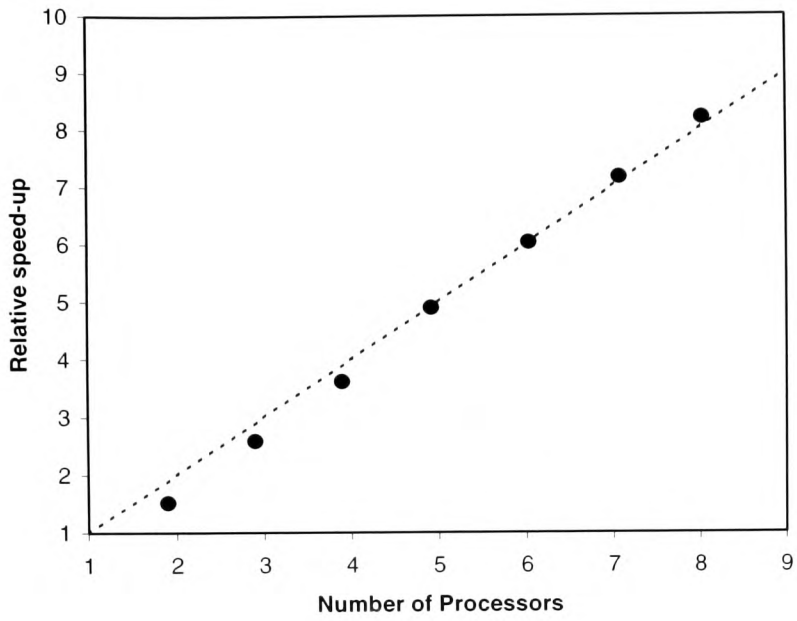


Figure 9.14 The relative speed-up for the parallel DGA optimization of the "bump function" (equation 6.44) running on a CRAY-T3D using different processor domain sizes, averaged over five runs.

9.6. CONCLUSIONS

The combined results of Chapter 8 and those above illustrate the flexibility of the FRAMEWORK model. The model has been shown to predict CO_2 fluxes accurately in single-layer big-leaf, or multi-layer configuration for two widely different plant types. The canopy structure, photosynthesis and radiation objects used for each plant type were chosen to represent accurately the processes of a millet or black spruce canopy. Although markedly different, these objects were easily integrated with the rest of the SVAT. The optimization routines integrated within FRAMEWORK, have been successfully used both to calibrate the modelled diurnal CO_2 flux predictions through the tuning of canopy nitrogen concentration, and to constrain the carbon assimilation of the canopy against remotely sensed data over a much longer temporal scale.

Chapter 10

Summary and Conclusions

10.1. SUMMARY

The response of vegetation to changes in climatic conditions has implications both for the modelling of future climate change through feedback effects between vegetation and the atmosphere, and for the future of human civilisation through changes rainfall patterns, crop production and land use. This study examined the use of remote sensing techniques to parameterize and validate vegetation models. Vegetation models are a vital part of research into climate change. They provide a means of scaling up measurements made at small spatial and temporal scales to larger scales relevant to the problems of global environmental change, and allow predictions to be made regarding the behaviour of vegetation under future climate change scenarios.

There are several reasons why the use of remote sensing techniques in SVAT modelling is desirable:

- it would allow the application of models over wide areas of the Earth's surface,
- it would allow frequent repeat measurements and
- it would allow significant cost savings by avoiding the need for field measurements in remote or dangerous parts of the world.

Chapter 2 considered previous attempts to use remote sensing in the context of vegetation modelling. The vast majority of studies have concerned the determination of model parameters from remotely sensed data. Unfortunately, determining the detailed parameter sets typically required by SVAT models using remote sensing techniques is extremely difficult for several reasons:

- the physical distance between the sensor and target,
- the heterogeneity of the Earth's surface,
- the complex three-dimensional nature of vegetation canopies and
- the rapidly changing nature of the land-atmosphere system.

Chapter 2 discussed the three main categories of remote sensing, optical, passive microwave and active microwave remote sensing. Unfortunately, the poor spatial resolution obtained by passive microwave sensors precludes the use of this technique at sub-regional scales. Active microwave (radar) techniques

are far more promising for stand-scale studies because of the much higher spatial resolutions that may be obtained. However, difficulties exist both in interpreting the backscatter signal and in processing the very large quantity of raw data produced by these sensors. The design of future sensors will be improved through the incorporation of multi-frequency instruments and development of backscatter models such as that of Sun *et al.* (1991). It is likely that such models will also improve our understanding of backscattering from vegetation canopies, enabling determination of several parameters useful to vegetation modellers, including biomass, surface roughness and soil surface water content. To date, several studies (described in Chapter 2) have obtained good correlations between biomass, LAI and the backscatter signal from the surface (Ranson Saatchi and Sun 1995; Ranson *et al* 1996) but the limited availability of microwave datasets makes the use of this technique unattractive for vegetation modellers at the present time.

Although optical wavelength remote sensing images require complex atmospheric corrections and cloud pixel decontamination, these techniques have been used in the vast majority of studies to date. Most of these studies have involved the construction of empirical relationships between vegetation properties and vegetation indices (VIs). Unfortunately, VIs are sensitive to changes in background reflectance caused by changes in climatic conditions such as rainfall, and become insensitive to changes in canopy properties at relative small LAIs through signal saturation.

In the last few years a better understanding of the anisotropic reflective properties of vegetation surfaces, characterised by the bidirectional reflectance distribution function (BRDF) of the canopy, has led to the development of multi-directional sensors. BRDF data contain much more information about the surface than VI measurements, but complex BRDF models are required in order to interpret multi-directional data. Chapter 2 concluded that although BRDF remote sensing techniques are currently the best choice for many applications, at best, the SVAT parameters obtainable from BRDF data are limited to LAI, biomass and APAR.

The difficult problem of modelling the anisotropic reflectance from vegetation was discussed in Chapter 3, where four BRDF models were presented. Because of the diverse range of different canopy structures exhibited by different vegetation types, BRDF models are usually tailored to a particular canopy type. However, the complexity of the calculations involved means it is necessary to compromise model accuracy through the introduction of simplifications and assumptions, in order to improve computation times.

Deriving canopy parameters from a remotely sensed BRDF signal requires BRDF model inversion. The complexity of BRDF models means that a large number of measurements are required for a successful inversion. However, the number of directional measurements available from remote sensing

instruments is limited. The solution to this problem proposed in this study was to use spectral information in addition to directional information. This solution requires a BDRF model with parameters that are independent of wavelength. Consequently, it was necessary to couple models of leaf and soil optical properties with BRDF models, in order to express the spectral reflectance in terms of wavelength independent parameters. Chapter 4 discussed models of the optical properties of leaves (the PROSPECT model) and soils (the SOILSPECT model) used in this study. It was found that the PROSPECT model is able to predict accurately the spectral reflectance of leaves using only five, wavelength-independent parameters. Excellent results were obtained by inverting the model using measured data for healthy leaves for species as diverse as millet and black spruce.

Few models of soil spectral reflectance have been developed that have wavelength independent parameters. The SOILSPECT model described in Chapter 4 represents the bi-directional reflectance of bare soils quite well but contains one parameter, the single scattering albedo, ω, that is not wavelength independent. The single scattering albedo governs the spectral behaviour of soil reflectance and is dependent on both soil type and moisture status. In this study, a look-up table approach was adopted to overcome the problem of the dependence of ω on wavelength and soil type. This was achieved by tabulating values of ω for a given spectral range for a number of different soil types. The problem of dependence on soil water content was overcome by modifying SOILSPECT to include a term representing the absorption properties of soil-bound water, derived from studies by Lekner and Dorf (1988) and Mauser (1996). These modifications resulted in the introduction of one extra input parameter into SOILSPECT, namely the water layer thickness, but made SOILSPECT effectively wavelength independent. By coupling the versions of the PROSPECT and SOILSPECT models developed in Chapter 4 with a BRDF model, a coupled model may be created that contains only wavelength independent parameters and is therefore invertible against multi-directional, multi-spectral data.

In almost all previous studies an inversion-parameterization strategy has been used to derive vegetation model parameters from BRDF data by inverting BRDF models. The success of these studies has been limited. Chapter 5 described a new parameterization-optimization strategy in which the output from a linked SVAT-scene model is compared with remotely sensed data over time, and the model constrained to achieve the best fit. In Chapter 2 it was concluded that with current technology remotely-sensed data will not replace the need for ground-based measurements: the parameterization-optimization strategy uses remotely sensed data to interpolate between ground-based measurements, enabling more accurate estimates of growth and yield.

The inversion or optimization of BRDF models cannot usually be performed analytically. Instead, trial and error techniques must be used. In Chapter 6 four optimization techniques, downhill simplex, direction set, simulated annealing and genetic algorithms were described and their performance

compared. Whilst effective on simpler problems, the downhill simplex and direction set methods struggled to find global optima in the presence of local optima. In contrast the two stochastic methods, simulated annealing and genetic algorithms, were able to find global optima under extremely difficult conditions. Unfortunately, these stochastic techniques are very much more intensive computationally. Consequently their use is restricted by the availability of computing resources.

Studies of the applications of the four optimizations techniques to the inversion-parameterization and parameterization-optimization techniques were carried out in Chapters 7 and 8 respectively. It was found that the Downhill Simplex and Direction Set techniques gave poor results. However, the stochastic optimization techniques simulated annealing (SA) and genetic algorithms (GA) achieved good results for the inversion of the PROSAIL+ model using a realistic dataset of directional and spectral reflectance measurements for a fully closed soybean canopy. SA and GA also retrieved LAI with good accuracy for an open soybean canopy through the inversion of a row-modified PROSAIL+ model. It was found that the performance of the stochastic techniques was dependent on the choice of control parameters used. The selection of the optimum set is largely a process of trial and error and tends to be application specific. Genetic algorithms were found to be less sensitive to these parameters than simulated annealing. However, once a tuned set of parameters has been obtained the results of SA are comparable to those of the GA.

Chapter 8 described the application of optimization techniques and the parameterization-optimization strategy to SVAT models. Previously SVAT models have been designed for a specific vegetation type. This has resulted in models that contain approximations and assumptions appropriate for a single vegetation type. Moreover, program design has meant that it has not been possible to easily inter-change sub-processes within the model. In Chapter 8 the development of a new SVAT model, FRAMEWORK, was described. Through careful design and the use of object oriented programming techniques FRAMEWORK is able to combine remotely sensed data from a number of sources in a synergistic way. FRAMEWORK is also very flexible both in terms of its representation of the canopy, and in terms of the model time step that may be used. In this respect FRAMEWORK may be regarded as a *software scaffold* upon which different modules may be built and readily interchanged.

In order to facilitate the application of either the inversion-parameterization or the valiadtion-optimization strategies a suite of optimization techniques was incorporated within FRAMEWORK. Optimization proceeds through the minimization of a merit function that may be constructed from any combination of model parameters, from any module within FRAMEWORK: this allows complete flexibility when optimizing the model. The flexible structure of FRAMEWORK and integrated optimization module allow the model to be confronted with remotely sensed data from a number of sources. In this study a bi-directional reflectance module was used, both to derive absorbed radiation and to predict canopy anisotropic reflectance.

Chapter 8 described the validation of FRAMEWORK using data for millet from the HAPEX-Sahel experiment in Niger. Good agreements were obtained between measurements of canopy CO_2 fluxes and growth, and model predictions. The differences between modelled and measured values were small and could be attributed to uncertainties in canopy structural parameters, canopy heterogeneity and measurement and optimization noise. It was concluded in Chapter 8 that these results demonstrated that the combination of stochastic optimization techniques and the validation-optimization technique is a powerful tool for vegetation modelling.

Although the effectiveness of stochastic optimization techniques seems clear provided sufficient computing power is available, the time taken for these techniques to find optima for complex SVATs, with multiple canopy layers and short time-steps is prohibitive with conventional computers. The ability of high performance computing (HPC) techniques to overcome this problem was investigated by modifying FRAMEWORK to run on parallel computer architectures using the MPI language. The results, described in Chapter 9, indicate that parallel computing techniques can provide a successful way of improving the performance of complex SVAT models. Chapter 9 illustrated that the use of HPC with genetic algorithms also enables complex SVATS to be optimized by constraining them against remotely sensed data, in a manner that would not be possible using conventional computing methods.

10.2. SUGGESTIONS FOR FUTURE WORK

Throughout this study it was apparent that the lack of suitable datasets was a limiting factor for research in this field. Very few stand-scale BRDF measurements have been made over crop canopies, and of those datasets that exist, few contain multi-spectral data in addition to multi-directional data. Moreover, BRDF measurements have usually been made in isolation, i.e. without corresponding canopy physical and biochemical measurements.

Detailed validations of the application of inversion techniques described in Chapter 7 and Chapter 8 require not only directional reflectance and physiological measurements for vegetation canopies, but also that these measurements be repeated several times during the growing season. The majority of recent remote sensing research has involved studies of forests. Although of great importance globally, forest canopies are both more difficult to measure physically and more difficult to model technically than crop canopies. Therefore, there is clearly a need for more crop canopy datasets comprising both ground-based physiological and remotely sensed measurements. Such datasets would be difficult and expensive to measure and would require co-ordination between different groups of scientists but would provide a much needed source of data.

Although for most BRDF models analytical inversion is either impossible or impractical, for one recently developed class of BRDF models, known as linear kernel-driven models, analytical inversion is possible (Lewis 1995). These models attempt to describe the BRDF as a superposition of a set of semi-emprical kernels that describe basic BRDF shapes, with weights assigned to each kernel chosen to optimize the fit between predictions and observations. Typically, kernels are based on approximations to a radiative transfer scenario for a horizontally homogeneous canopy, or approximations in various geometrical-optical representations of scattering from a scene containing discrete objects (Wanner, Li and Strahler 1995a).

If a model can be simplified to a point where terms containing geometric expression are isolated from the rest of the model parameters, then these geometric terms can be used as *kernels*, and following the scheme proposed by Roujean, Leroy and Deschamps (1992) the BRDF function may be written as:

$$R = f_{iso} + f_{surf} k_{surf} + f_{vol} k_{vol} \qquad (10.1)$$

where R is the canopy reflectance, k is a kernel describing the surface or volume scattering and f are the respective weights, where f_{iso} is a Lambertian constant. Semi-empirical, kernel-driven models have a number of advantages over physical and purely empirical models: they are must faster to compute than physical models, they require only a limited number of observations to be invertible, and, if the superposition of kernels is linear, the model may be inverted analytically by solving a set of linear equations for the minimum of an error function (Lewis 1995).

Kernel driven models have been successfully developed for use in albedo studies (Wanner, Li and Strahler 1995b). Unfortunately, the assumptions, simplifications and approximations made when deriving the kernels means that the parameters obtained by inverting the models have no direct meaning in terms of individual biophysical or structural parameters of the vegetation target. To date, there have been no attempts to modify the kernel-based approach in an attempt to derive vegetation parameters useful for SVAT modelling. However, because of the ability to invert these models analytical studies in this field are likely to become attractive to researchers in the near future.

One rôle of complex, multi-component, models is to examine the limitations of simpler models, such as big leaf models, by quantifying the errors resulting from the assumption of canopy homogeneity. The flexible structure of the FRAMEWORK model enables different representation of vegetation function to be rapidly incorporated into the model. This allows the implications of simplifying assumptions implicit in any representation to be investigated. In this way, the output of complex models may be used to infer modifications to simple models to introduce more realistic representations of function without greatly increasing the computational complexity of the model. FRAMEWORK provides a useful tool for future studies in this area, which should, in turn, lead to improvements in our understanding of canopy processes and the design of vegetation models.

One approach to dealing with the computational complexity of multi-component models is to introduce new computing techniques. In Chapter 9 of this study the P-FRAME model was successfully derived from FRAMEWORK. P-FRAME is able to utilise parallel processing techniques by distributing calculations across a number of separate processors. Three aspects of FRAMEWORK were identified as offering potential performance gains through parallelism: (i) patch-landscapes, (ii) multiple component vegetation structural configurations, and (iii) genetic algorithm optimization of the model. Even though the time available for P-FRAME code development was limited, significant speed-ups were achieved in all three of the above cases, demonstrating that high performance computing techniques could play a significant rôle in future vegetation modelling strategies. Further development of parallel processing SVAT models would enable more complex simulations to be performed and would facilitate the coupling of SVAT and atmospheric models to examine the feedback effects between vegetation and the atmosphere. As part of the development of P-FRAME, Chapter 9 also described the implementation of a parallel genetic algorithm optimization technique. Parallel GAs have great potential as powerful optimization tools, but their properties are still not fully understood. Further development of parallel SVAT modelling techniques would provide an excellent opportunity for further study of parallel GAs.

Whilst the radiation routines used within P-FRAME gave excellent APAR predictions during this study, they do not enable complex scattering of bi-directional reflectance to be modelled. An important improvement therefore, would be to incorporate a more sophisticated radiation transfer model, such as the DART model (Gastellu-Etchegorry *et al* 1995), into FRAMEWORK. This model simulates radiative transfer in heterogeneous three-dimensional scenes using the discrete ordinate method (Kimes and Kirchner 1982). The scene is divided into rectangular cell matrices, with dimensions along the x,y,z axes that are not necessarily equal. A phase function is computed for each cell using its optical and geometrical parameters. Both direct and diffuse radiation from the atmosphere may be aniostropic. The model processes the interactions of each individual source vector with all the cells it encounters as it propagates through the scene. Direct solar source vectors give rise to secondary source vectors via the cell scattering functions. Diffuse atmospheric source vectors and the secondary source vectors resulting from within cell scattering give rise to tertiary vectors and so on until an energetic equilibrium is reached. This detailed treatment of radiative transfer, although computationally intensive, would enable accurate simulations of vegetation canopies and consequently better SVAT model predictions. The structure of DART is such that it should be possible to decompose the model to run on parallel hardware with ease, and when combined with P-FRAME should prove a powerful modelling tool.

References

Abuelgasim, A.A. and Strahler, A.H. (1994)
Modelling bidirectional reflectance measurements collected by the advanced soild-state array spectroradiometer (ASAS) over Oregon transect conifer forests.
Rem. Sens. Env. 47:261-275

Acock, B. and Reddy, V.R. (1997)
Designing an object-oriented structure for crop models.
Ecological Modelleing 94:33-44

Allen, W.A. (1973).
Transmission of isotropic light across a dielectric surface in two and three dimensions.
Jour. Opt. Soc. Am. 63:664-666

Allen, W.A. and Richardson, A.J. (1968).
Interaction of light with a canopy.
Jour. Opt. Soc. Am. 58:1023-1028

Allen, W.A., Gausman, H.W., and Richardson, A.J. (1970).
Mean effective optical constants of cotton leaves.
Jour. Opt. Soc. Am. 60:542-547

Allen, W.A., Gausman, H.W., Richardson, A.J. and Thomas, J.R. (1969).
Interaction of light with a compact plant leaf.
Jour. Opt. Soc. Am. 59:1376-1379

Anderberg, M.R. (1975)
Cluster analysis for applications.
Academic Press

Anderson, M.C. (1964)
Studies of the woodland climate. 1 The photographic computation of light conditions.
J. Ecol. 52:27-41

Angstrom, A (1925)
The albedo of various surfaces of ground.
Geogr. Ann. 7:323-325

Aphalo, P.J. and Jarvis, P.G. (1993)
An analysis of Ball's empirical model of stomatal conductance.
Ann. Bot. 72:321-327

Asrar, G.M. (1989)
Theory and applications of optical remote sensing.
Wiley

Asrar, G.M., Fuchs, E.T., Kanemasu, E.T. and Hatfield, J.L. (1984)
Estimating absorbed photosynthetically active radiation and leaf area index from spectral reflectance in wheat.
Agron. J. 76:300-306

Azam-Ali, S.N., Gregory, P.J. and Monteith, J.L. (1984)
Effects of planting density on water use and productivity of pearl millet grown on stored water. II. Water use, light interception and dry matter production.
Exp, Agric. 20:215-224

Baret, F. Andrieu, B. and Guyot, G. (1988).
A simple model of leaf optical properties in the visible and near-infrared: application to the analysis of spectral shifts determinism.
In: *Applications of chlorophyll flourescence* (H. Lichtenthaler, Ed.) Kluwer Academic. pp 345-351

Baret, F., Guyot, G. and Major, D.J. (1989)
TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation
In: *Proc. 12th Can. Symp.. Rem. Sens./ 1989 Int. Geosci. Rem. Sens. Symp., Vancouve.* pp 1355-1358

Barnola, J.M., Raynaud, D., Korotkevitch, Y.S. and Lorius, C. (1987)
Vostock ice core: a 160,000 year record of atmospheric CO_2.
Nature 329:408-414

Baumgardner, M.F., Silva, L.F., Biehl, L.L. and Stoner, E.R. (1985)
Reflectance properties of soils.
Adv. Agron. 38:1-44

Beaudoin, A., Le Toan, T. and Gwyn, Q.H.J. (1990)
SAR observations and modelling of the C-band backscatter variability due to multiscale geometry and soil moisture.
IEEE Trans. Geosci. Rem. Sens. 28:886-895

Berry, J.A. and Farquhar, G.D. (1978)
The CO_2 concentrating function of C_4 photosynthesis: a biochemical model.
In: Proceeding of the 4th international congress on photosynthesis. (eds: D. Hall, J. Coombs and T. Goodwin)
Biochemical Society, London.

Björkman, O. (1981a)
The response of photosynthesis to temperature.
In: *Plants and their atmospheric environment.*(eds: J. Grace, E.D. Ford and P.G. Jarvis)
Blackwell Scientific, Oxford.

Björkman, O. (1981a)
Responses to different quantum flux densities.
In: *Encyclopaedia of plant physiology*, Vol. 12A. *Physiological plant ecologyl, responses to the physical environment.* (Eds O.L. Lange, P.S. Nobel, C.B. Osmond and H. Ziegler.)
Springer Verlag, Berlin.

Björkman, O., Boardman, N.K., Anderson, J.M., Throne, S.W., Goodchild, D.J.
and Pyliotis , N.A. (1972)
Effect of light intensity during growth of *Atriplex patula* on the capacity of photosynthetic reactions, chloroplast components and structure.
Carnegie institute of washington year book. 71:115-135

Booch, G. (1991)
Object-Oriented Design With Applications.
Benjamin Cummings.

Booker, L. (1987)
Improving search in genetic algorithms.
In: *Genetic algorithms and simulated annealing.* (Ed. L Davis)
Morgan Kaufmann

Boyer, M., Miller, J., Belanger, M., Hare, E. and Wu, J. (1988)
Senescence and spectral reflection in leaves of northern pink oak (*Quercus palustris Muench.*)
Rem. Sens. Env. 25:71-87

Brown P.S. and Pandolfo, J.P. (1969)
An equivalent obstacle model for the computation of radiation flux in obstructed layers.
Agric. Met. 6:407-421

Calvet, J-C., Wigneron, J-P., Mougin, E, Kerr, Y.H. and Brito, J.L.S. (1992)
Plant water content and temperature of the Amazon forest from satellite microwave radiometry.
(*Personal communication*)

Camilo, P.and Schmugge, T.J. (1983)
Estimating soil moisture storage in the root zone from surface measurements.
Soil Sci. 135:245-264

Campbell, G.S. (1990)
Derivation of an angle density function for canopies with ellipsoidal angle distributions.
Agric. And Forest Met. 49:173-176

Campbell, G.S. (1986)
Extinction coefficients for radiation in plant canopies calculated using an ellipsoidal inclination angle distribution.
Agric. And Forest Met. 36:317-321

Campbell, G.S. and Norman, J.M. (1989)
The description and measurement of plant structure.
In: Plant canopies: their growth, form and function. (eds: G. Russell, B. Marshall and P.G. Jarvis.)

Cannell, M.G.R., Milne, R., Sheppard, L.J. and Unsworth, M.H. (1987)
Radiation interception and productivity of willow.
J. Appl. Ecol. 24:261-268

Chandrasekhar, S. (1950)
Radiative Transfer
Clarendon Press

Charles-Edwards, D.A. (1982)
Physiological determinants of crop growth.
Academic Press.

Chason, J.W., Baldocchi, D.D., Huston, M.A. (1991)
A comparison of direct and indirect methods for estimating forest canopy leaf area.
Agric. For. Met. 57:107-128

Choudhury, B.J. (1987)
Relationships between vegetation indices, radiation absorption and net photosynthesis evaluated by a sensitivity analysis.
Rem. Sens. Env. 22:209-233

Crist, E.P. and Ciccone, R.C. (1984)
A physically-based transformation of Thematic Mapper data – the TM tasseled cap.
IEEE Trans. Geosci. Rem. Sens. 22:256-263

Clevers, J.G.P.W. and Verhoef, W. (1991)
Modelling and synergistic use of optical and microwave remote sensing.
Report 2: LAI estimation from canopy reflectance and WDVI: a sensitivity analysis with the SAIL model.
BRCS Report 90-99.

Connolly, D. (1992)
General purpose simulated annealing.
J. Opt. Res. Soc. 43:495-505

Collatz, G.J., Ribas-Carbo, M. and Berry, J.A. (1992)
Coupled photosynthesis-stomatal conductance model for leaves of C_4 plants.
Aust. J. Plant. Physiol. 19:519-538

Curran P.J. (1981)
Multispectral remote sensing for estimating biomass and productivity.
In: Plants and the daylight spectrum. (ed: H Smith.)
New York Academic Press, pp. 65-99

Curran P.J., Dungan, J.L., MacIer, B.A. and Plummer, S.E. (1991)
The effect of red leaf pigment on the relationship between red egde and chlorophyll concentration.
Rem. Sens. Environ. 35:69-76

Demetriades-Shah, T.H., Fuchs, M., Kanemasu, E.T. and Flitcroft, I.D. (1992)
A note of caution concerning the relationship between cumulated intercepted solar radiation and crop growth.
Agric. For. Met. 58:193-207

Demetriades-Shah, T.H., Fuchs, M., Kanemasu, E.T. and Flitcroft, I.D. (1994)
Further discussions on the relationship between cumulated intercepted solar radiation and crop growth.
Agric. For. Met .68:231-242

Dewar, R.C. (1995)
The correlation between plant growth and intercepted radiation: an interpretation in terms of optimal plant nitrogen content.
Pers. Comm. (Submitted to Annals of Botany)

Dobson, M.C., Ulaby, F.T., Le Toan, T., Beaudoin, A., Kasischke, E.S. and Christensen, N. (1992).
Dependence of radar backscatter on coniferous forest biomass.
IEEE Trans. Geosci. Rem. Sens. 30:412-415

Eamus, D. and Jarvis, P.G. (1989)
The direct effect of increases in global CO_2 concentration on temperate trees and forests (natural and commercial).
Advances in Ecol. Research 19:1-55

Evans, J.R. (1988)
Acclimation by thylakoid membranes to growth irradiance and the partitioning of nitrogen between soluble and thylakoid protiens.
In: *Ecology of photosynthesis in Sun and Shade.* (eds. J.R. Evans, S. von Cammerer and W.W. Adams III).
CSIRO, Australia

Evans, J.R. (1987)
The dependence of quantum yield on wavelength and growth irradiance.
Australian Journal of Plant Physiology 14:69-79

Evans, J.R. (1988) and Farquhar, G.D.. (1991)
Modelling canopy photosynthesis from the biochemistry of the C_3 chloroplast.
In: *Modelling crop photosynthesis – from biochemistry to canopy.(Crop science society of America, No.19).* (eds. K.J. Boote and R.S. Loomins)
Crop society of America.

Farquhar, G.D. and von Caemmerer, S. (1982)
Modelling of photosynthetic response to environmental conditions.
In: *Encyclopaedia of plant physiology*, Vol. 12B. *Water relations and carbon assilimation.* (eds: O.L. Lange, P.S. Nobel, C.B. Osmond and H. Ziegler.)
Springer Verlag, Berlin.

Farquhar, G.D., von Caemmerer, S. and Berry, J.A. (1980)
A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species.
Planta. 149:78-90

Fassnacht, K.S., Gower, S.T., Norman, J.M. and McMurtrie, R.E. (1994)
A comparison of optical and direct methods for estimating foliage surface area index in forests.
Agric. For. Met. 71:183-207

Field, C. (1983)
Allocation of leaf nitrogen for the maximization or carbon gain: leaf age is a control on the allocation program.
Oecologia. 56:341-347

Fogel, D.B. (1993)
Applying evolutionary programming to selected travelling salesman problems.
Cybernetics and Systems 24:27-36

Forrest, S. and Mitchell, M. (1995)
Relative building block fitness and the building block hypothesis.
In: Foundations of genetic algorithms. (ed: L.D. Whitley.)
Morgan Kaufman

Gastellu-Etchegorry, J.P., Demarez, V., Pinel, V. and Zagolski, F. (1995)
Modelling radiative transfer in heterogeneous £D vegetation canopies.
Rem. Sens. Env. 52:163-172

Gausman, H.W. and Allen, W.A. (1973).
Optical parameters of leaves of 30 plant species.
Plant. Physiol. 52:57-62

Goel, N.S. (1987)
Models of vegetation canopy reflectance and their use in estimation of biophysical parameters from reflectance data.
Rem. Sens. Reviews 3:1-212

Goel, N.S. and Grier, T. (1988).
Estimation of canopy parameters for inhomogeneous vegetation canopies from reflectance data. III TRIM: a model for radiative transfer in heterogeneous three-dimensional canopies.
Remote Sens. Environ. 25:255-293

Goel, N.S. and Grier, T. (1986a).
Estimation of canopy parameters for inhomogeneous vegetation canopies from reflectance data. I Two-dimensional row canopy.
Int. J. Rem. Sens. 7:665-681

Goel, N.S. and Grier, T. (1986b).
Estimation of canopy parameters for inhomogeneous vegetation canopies from reflectance data. II Estimation of leaf area index and percentage of ground cover for row canopies.
Int. J. Rem. Sens. 7:1263-1286

Goel, N.S. and Strebel, D.E.(1984).
Simple beta distribution representation of leaf orientation in vegetation canopies.
Agronomy Journal. 76:800-802

Goel, N.S., Strebel, D.E. and Thompson, R.L. (1985).
Optimal solar/viewing geometry for an accurate estimation of leaf area index and leaf angle distribution from bidirectional reflectance data.
Int. J. Rem. Sens. 6:1493-1520

Goel, N.S., Strebel, D.E. and Thompson, R.L. (1984).
Inversion of vegetation canopy reflectance models for estimating agronomic variables. II Use of angle transforms and error analysis as illustrated by Suits' model.
Remote Sens. Environ. 14:77-111

Goel, N.S., and Thompson, R.L. (1984a).
Inversion of vegetation canopy reflectance models for estimating agronomic variables. III Estimation using only canopy reflectance data as illustrated by Suits' model.
Remote Sens. Environ. 15:223-236

Goel, N.S., and Thompson, R.L. (1984b).
Inversion of vegetation canopy reflectance models for estimating agronomic variables. IV Total inversion of the SAIL model.
Remote Sens. Environ. 15:237-253

Goel, N.S., and Thompson, R.L. (1984c).
Inversion of vegetation canopy reflectance models for estimating agronomic variables. V Estimation of leaf area index and average leaf angle using measured canopy reflectances.
Remote Sens. Environ. 16:69-85

Goldberg, D. (1987)
Simple genetic algorithms and the minimal, deceptive problem.
In: *Genetic algorithms and simulated annealing.* (Ed. L. Davis)
Pitman

Goldberg, D. and Deb, K. (1991)
A comparative analysis of selection schemes used in genetic algorithms.
In: *Foundations of genetic algorithms.* (Eds G. Rawlins).
Morgan-Kaufmann.

Goudriaan, J. and Monteith, J.L. (1990)
Amathematical function for crop growth based on light interception and leaf area expansion.
Ann. Bot.

Goutorbe, J. P., Lebel, T. and Tinga, A. et al. (1994)
HAPEX-Sahel: a large-scale study of land-atmosphere interactions in the semi-arid tropics.
Annales Geophysicae. 12:53-64.

Goward, S.N. and Huemmrich, K.F. (1992).
Vegetation canopy PAR absorptance and the normalized vegetation index: an assessment using the SAIL model.
Rem. Sens. Env. 39:119-140

Gower, S. and Norman, J.M. (1991)
Rapid estimation of leaf area index in conifer and broad leaf plantations.
Ecology 72:1896-1900

Grace, J. (1983)
Plant-Atmosphere Relationships.
Chapman and Hall, London.

Grace, J., Lloyd, J., McIntyre, J. Miranda, A.,C., Miranda, H., J., Wright, I.R. and Gash, J. (1995a)
Fluxes of water vapour and carbon dioxide over an undisturbed tropical rain forest in South-West Amazonia.
Global Change Biology 1:1-12

Grace, J., Lloyd, J., McIntyre, J. Miranda, A.,C., Meir, P., Miranda, H., Nobre, C., Moncreiff, J.B., Massheder, J., Malhi, Y., Wright, I.R. and Gash, J. (1995)
Carbon dioxide uptake by an undisturbed tropical rain forest in South-West Amazonia.
Science 270:778-780

Grant, L. (1987).
Diffuse and specular characteristics of leaf reflectance.
Rem. Sens. Env. 22:309-382

Hall F.G., Huemmrich K.F., and Goward S.N. (1990)
Use of narrow band spectra to estimate the fraction of absorbed photosynthetically active radiation.
Remote Sens. Environ. 32:47-54

Heute, A.R. (1987)
Soil and sun angle interactions on partial canopy spectra.
Int. J. Rem. Sens. 8:1307-1317

Heute, A.R. (1988)
Soil adjusted vegetation index.
Rem. Sens. Env. 25:295-309

Heute, A.R. and Jackson, R.D. (1988)
Soil and atmosphere influences on the spectra of partial canopies.
Rem. Sens. Env. 25:89-105

Holinger, D.Y., Kelliher, F.M., Byres, J.N., Hunt, J.E., McSevney, T.M. and Weir, P.L. (1994)
Carbon dioxide exchange between an undisturbed old-growth temperate forest and the atmosphere.
Ecology 75:134-150

Holland, J. (1975)
Adaptation in natural and artificial systems.
University of Michigan Press

Hunt, E.R. Jnr and Rock, B.N. (1989)
Detection of changes in leaf water content using near-infrared and middle0infrared reflectances.
Rem. Sens. Env. 30:43-54

Ishimaru, A. (1978a)
Wave propagation and scattering in random media, Vol 1. Single scattering and transport theory.
Academic Press

Ishimaru, A. (1978b)
Wave propagation and scattering in random media, Vol 2. Multiple scattering, turbulence, rough surfaces and remote sensing.
Academic Press

Jackson, R.D., Moran, M.S., Gay, W.L. and Raymond, L.H. (1987)
Evaluating evaporation from field crops using airborne radiometry and ground-based meteorological data.
*Irrig. Sci.*8:81-90

Jackson, R.D., Slater, P.N. and Pinter, P.J. (1983)
Discrimination of growth and water stress in wheat by various vegetation indices through clear and turbid atmospheres.
Rem. Sens. Env. 13:187-208

Jackson, T.J. and Schmugge, T.J. (1991).
Vegetation effects on microwave emission from soils.
Rem.Sens. Env. 36:203-212.

Jacquemoud, S. (1993).
Inversion of the PROSPECT and SAIL canopy reflectance models from AVRIS equivalent spectra: theoretical study.
Rem. Sens. Env. 44:281-292

Jacquemoud, S. and Baret, F. (1990).
PROSPECT - a model of leaf optical properties.
Rem. Sens. Env. 34:75-91

Jacquemoud, S., Baret, F. Andrieu, B., Danson, F.M. and Jaggard, K. (1995).
Extraction of vegetation biophysical parameters by inversion of the PROSPECT and SAIL models on sugar beet canopy reflectance data. Application to TM and AVRIS sensors.
Rem. Sens. Env. 52:163-172

Jacquemoud, S., Baret, F. and Hanocq, J.F. (1992).
Modelling spectral and bidirectional soil reflectance.
Rem. Sens. Env. 41:123-132

Jarvis, P.G. (1976)
The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field.
Phil. Trans. Royal. Soc. Lond. Series B, 272,:593-610
Jarvis, P.G. (1993)
Prospects for bottom-up models.
In: *Scaling physiological processes: leaf to globe.* (Eds: J. Ehleringer and C. Field)
Academic Press, New York.

Jarvis, P.G. and McNaughton, K.G. (1986).
Stomatal control of transpiration: scaling up from leaf to region.
Advances Ecol. Res. 15:1-49

Jarvis, P.G. and Leverentz, J.W. (1983)
Productivity of temperate, deciduous and evergreen forests.
In: *Encyclopedia of plant physiology. Vol. 12 Physiological plant ecology IV.* (eds. O.L. Lange, P.S. Nobel, C.B. Osmond and H. Ziegler)
Springer-Verlag

Keane, A.J. (1995)
Genetic algorithm optimization of multi-peak problems: studies in convergence and robustness.
Artificial Intelligence in Engineering 9:75-83

Keeling, C.D., Bacastow, R.B., Carter, A.F., Piper, S.C., Whorf, T.P., Heimann, M., Mook, W.G. and Roeloffzen, H. (1989)
A three dimensional model of atmospheric CO_2 transport based on observed winds: 1. Analysis of observational data.
In: *Aspects of climate variability in the Pacific and Western Americas.* Geophysical monograph 55; D.H. Peterson ed.)

Kiang, R.K. (1982)
Atmospheric effects on TM measurements: Characterisation and compariso with effects on MSS.
IEEE Trans. Geosci. Rem. Sens. 20:365-370

Kimes, D.S. (1984)
Modelling the directional reflectance from complete homogeneous vegetation canopies with various leaf orientation distributions.
J. Opt. Soc. America 22:124-132

Kimes, D.S. and Kirchner, J.A. (1982)
A radiative transfer model for heterogeneous 3D scenes.
Appl. Optics 21:4119-4129

Kimes, D.S., Markham, B.L., Tucker, C.J. and McMurtrey, J.E. (1981)
Temporal relationships between spectral response and agronomic variables of a corn canopy.
Rem. Sens. Env. 11:401-411

Kimes, D.S., Newcomb, W.W., Tucker, C.J., Zonneveld, I.S., van Wijngaarden, W. and de Leeuw, J. (1985).
Directional reflectance factor distributions for cover types of northern Africa.
Rem. Sens. Env. 18:1-19

Kimes, D.S. and Sellers, P.J. (1985)
Inferring hemispherical reflectance of the earth's surface for global energy budgets from remotely sensed nadir or directional radiance values.
Rem. Sens. Env. 18:205-223

Kiniry, J.R. (1994)
A note of caution concerning the paper by Demetriades-Shah et al (1992).
Agric. For. Met. 68:229-230

Kirkpatrick, S. (1984).
Optimization by simulated annealing: quantitative studies.
J. Stat. Phys. 34:975-986.

Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983).
Optimization by simulated annealing.
Science 220:671-676

Knuth, D.E. (1981)
Semi-numerical algorithms: The art of computer programming, volume 2.
Addison-Wesley

Kruijt, B., Ongeri, S. and Jarvis, P.G. (1997)
Scaling of PAR absorption, photosynthesis and transpiration from leaves to canopy.
In: *Scaling up from cell to landscape*, SEB seminar series 63 (eds: P.R. van Gardingen, G.M. Foody and P.J. Curran).
Cambridge University Press

Kumar, M. and Monteith, J.L. (1981)
Remote sensing of crop growth.
In: Plants and the daylight spectrum (H smith, Ed.)
New York Academic Press, pp. 134-144

Landsberg, J.J., and James, G.B. (1971)
Wind profiles in plant canopies: studies on an analytical model.
J. App. Ecol. 8:729-741

Landsberg, J.J., and Jarvis, P.G. (1973)
A numerical investigation of the momentum balance of a spruce forest.
J. App. Ecol. 10:645-655

Landsberg, J.J., Prince, S.D., Jarvis, P.G., McMurtrie, R.E., Luxmoore, R. and Medlyn, B.E. (1995)
Energy conversion and use in forests: the analysis of forest production in terms of radiation utilization efficiency (ε).
In: *The use of remote sensing in the modelling of forest productivity at scales from the stand to the globe.* (eds. H.L. Gholz and K. Nakane)
Kluwer Academic Press.

Landsberg, J.J. and Wright, L.L. (1989)
Comparisons among Populus clones and intensive culture conditions, using an energy conversion model.
For. Ecol. Management 27:129-147

Lang, A.R.G., Xiang, Y. and Norman, J.M. (1985)
Crop structure and the penetration of direct sunlight.
Agric. For. Met. 35:83-101

Lekner, J. and Dorf, M.C. (1988)
Why some things are darker when wet.
App. Opt. 27:246-251

Leuning, R. and Moncrieff, J.B. (1990)
Eddy covariance CO_2 flux measurements using open and closed path CO_2 analysers: corrects for analyser water vapour sensitivity and damping of fluctuations in air sampling tubes.
Boundary Layer Met. 53:63-76

Leverentz, J.W., Falk, S., Pilström, C-M. and Samuelsson, G. (1990)
The effects of photoinhibition on the photosynthetic light-response curve of green plant cells (*Chlamydomonas reinhardtii*).
Planta. 182:161-168

Lewandowska, M., Hart, J.W. and Jarvis, P.G. (1977)
Photosynthetic electron transport in shoots of Sitka spruce from different levels in a forest canopy.
*Physiologia Planatarum.*41:124-128

Lewis, P. (1995)
The utility of kernel driven BRDF models in global BRDF and albedo studies.
Proc. Int. Geosci. Rem. Sens. Symp. 95:1186-1188

Li, X. and Strahler, A.H. (1985)
Geometric-optical modelling of a conifer forest canopy.
IEEE Trans. Geosci. Rem. Sens. 23:705-712

Li, X. and Strahler, A.H. (1986)
Geometric-optical bidirectional reflectance modelling of a conifer forest canopy.
IEEE Trans. Geosci. Rem. Sens. 24:906-919

Li, X. and Strahler, A.H. (1988)
Modelling the gap probability of a discontinuous vegetation canopy.
IEEE Trans. Geosci. Rem. Sens. 26:161:170

Li, X. and Strahler, A.H. (1992)
Geometric-optical bidirectional reflectance modelling of mutual shadowing effect of crown shape and mutual shadowing.
IEEE Trans. Geosci. Rem. Sens. 30:276-292

Liou, K.N. (1980)
An introduction to atmospheric radiation.
International Geophysics Series. Vol 26

Maas, S.J. (1988).
Using satellite data to improve model estimates of crop yield.
Agron. Journ. 80:655-662

Maas, S.J. and Dunlap, J.R. (1989).
Reflectance, transmittance and absorptance of light by normal, etiolated and albino corn leaves.
Agron. Journ. 81:105-110

McNaughton, K.G. and Jarvis, P.G. (1991)
Effects of spatial scale on stomatal control of transpiration.
Agric. For. Met. 54:279-301

Madgwick, H.A. and Broomfield, G.L. (1969)
The use of hemispherical photographs to assess the light climate in a forest.
J. Ecol. 57:537-542

Major, D.J., Schaalje, G.B., Wiegand, C. and Blad, B.L. (1992)
Accuracy and sensitivity analyses of SAIL model predicted reflectance of maize.
Rem. Sens. Env. 41:61-70

Mauser (1996)
The spectral reflectance of soil under varying moisture conditions.
Pers. Comm.

McDonald, K.C., Zimmermann, R., Way, J.B. and Oren, R. (1990).
Using MIMICS to model L-band multi-angle and multi-temporal backscatter from a walnut orchard.
IEEE Trans. Geosci. Rem. Sens. 28:477-491

McMurtrie, R.E. and Wang, Y-P. (1993)
Mathematical models of the photosynthetic response of tree stands to rising CO_2 concentrations and temperatures.
Plant Cell Env. 16:1-13

Medlyn, B.E. (1996)
Modelling of forest response to elevated CO_2 under nitrogen limitation.
PhD Thesis University of New South Wales.

Moncrieff, J.B., Monteny, B., Verhoef, A., Friborg, Th., Elbers, J., Kabat, P. de Bruin, H., Soegaard, H., Jarvis, P.G. and Taupin, J.D. (1997)
Spatial and temporal variations in net carbon fluc during HAPEX-Sahel.
J. Hydrology 188:563-588

Monteith, J.L. (1995)
A reinterpretation of stomatal responses to humidity.
Plant Cell Env. 18:357-364

Monteith, J.L. (1994)
Validity of the correlation between intercepted radiation and biomass.
Agric. For. Met. 68:213-220

Monteith, J.L. (1981)
Does light limit crop production?
In: Physiological processes limiting plant production. (ed: C.B. Johnson)
Butterworths, London.

Monteith, J.L. (1977).
Climate and the efficiency of crop production in Britain.
Phil. Trans. R. Soc. Lond. B. 281:277-294

Monteith, J.L. (1972)
Solar radiation and productivity in tropical ecosystems.
J. App. Ecol. 9:747-766

Monteith, J.L., Marshall, B., Saffell, R.A., Clark, D., Gallagher, J.N., Gregory, P.J., Ong, C.K., Squire, G.R. and Terry, A. (1983)
Environmental control of a glasshouse suite for crop physiology.
J. Exp. Bot. 24:309-321

Moneith, J.L. and Unsworth, M.H. (1990)
Principles of environmental physics. (2^{nd} ed.)
Edward Arnold

Muhlenbein, H. (1991)
Evolution in time and space – the parallel genetic algorithm.
In: *Foundations of genetic algorithms* (ed. G.J.E. Rawlins)
Morgan Kaufmann

Myneni, R.B., Asrar, G., Gerstl, S.A.W. (1990)
Radiative transfer in three dimensional leaf canopies.
Transport theory and statistical physics. 19:205-250

Neftel, A., Moor, E., Oeschger, H. and Stauffer, B. (1985)
Evidence from polar ice cores for the increase in atmospheric CO_2 in the past to centuries.
Nature 315:45-47

Nelder, J.A. and Mead, R. (1965)
A simplex method for function minimization.
Computer Journal 7:308-398

Nemani, R.R. and Running , S.W. (1989)
Testing a theoretical climate-soil-leaf area hydrologic equilibrium of forests using satellite data and ecosystem simulation.
Agric For. Met. 44:245-260

Nilson, T., Kuusk, A., Joffre, R., Lacaze, B., Methy, M. and Peterson, U. (1995)
Determination of forest ecosystem parameters by inversion of a forest reflectance model.
In: *Proc. Int. Colloq. Phototsysnthesis and Remote Sensing.* Montpellier. pp149-157

Norman, J.M. (1978)
Modelling the complete crop canopy.
In: *Modification of the aerial environment of crops.* (Eds. B.J. Barfield and J.F. Gerber)
ASAE Monograph, American Society of Agricultural Engineering.

Norman, J.M. and Campbell, G.S. (1989)
Canopy structure.
In Plant physiological ecology: field methods and instrumentation. (R.W. Pearcy J. Ehleringer, H. A. Mooney and P.W. Rundel eds)
Chapman & Hall, London. pp301-325

Norman, J.M., Garcia, R. and Verma, S.B. (1992)
Soil surface CO_2 fluxes and the carbon budget of grassland.
J. Geophys. Res. 97:18845-18543.

Norman, J.M. and Jarvis, P.G. (1975)
Photosynthesis in Sitke spruce (*Picea sitchensis* (Bong) Carr.) V. Radiation penetration theory and a test case.
J. Applied Ecology 12:839-878

Norman , J.M. and Welles, J.M. (1979)
Radiative transfer in an array of canopies.
Agronomy J.

Norman , J.M., Welles, J.M. and Walter, E.A. (1985)
Contrasts among bidirectional reflectance of leaves, canopies and soils.
IEEE Trans. Geosci. Rem. Sens. 23:659-668

Oberbauer, S.F., and Strain, B.R. (1986)
Effects of canopy position and irradiance on the leaf physiology and morphology of *Pentaclethra macroloba* (Mimosaseae).
American J. of Botany. 73:409-416

Oker-Blom, P. (1984)
Penumbral effects of within plant and between plant shading on radiation distribution and leaf photosynthesis: a Monte Carlo simulation.
Photosynthetica 18:522-528

Ong, C.K. and Squire, G.R. (1984)
Response to temperature of a stand of pearl millet. VII. Final number of spikelets and grains.
J. Exp. Bot. 34:309-321

Palmer, K.F. and Williams, D. (1974)
Optical properties of water in the near infrared
J. Opt. Soc. Am. 64:1107-1110

Park, S.K., Miller, K.W. and Stockmeyer, P.K. (1993)
Random number generators and the minimal standard.
Communications of the Association for Computing Machinery 36:108-110

Park, S.K. and Miller, K.W. (1989)
Random number generators and the minimal standard.
Communications of the Association for Computing Machinery 32:1023-1024

Park, S.K. and Miller, K.W. (1988)
Random number generators – good ones are hard to find.
Communications of the Association for Computing Machinery 31:1192-1201

Penning de Vries, F. W. T., Jansen, D. M., ten Berge, H. F. M. and Bakema, A. (1989)
Simulation of ecophysiological processes of growth in several annual crops.
Pudoc, Wageningen.

Peterson, D.L. and Running, S.W. (1989)
Applications in forest science and management
In: Theory and applications of optical remote sensing, Chapter 10 (G. Asrar, ed.).
Wiley pp. 429-473

Peteterson, D.L., Aber, J.D., Matson, P.A., Card, D.H., Swanberg, N., Wessman, C. and Spanner, M. (1988).
Remote sensing of forest canopy and leaf biochemical contents.
Rem. Sens. Env. 24:85-108

Phillips, P.C. (1993)
Peak shifts and polymorphism during phase three of Wright's shifting-balance process.
Evolution 47:1733-1743

Pierce, L.L and Congalton, R.G. (1988)
A methodology for mapping forest latent heat flux densities using remote sensing.
Rem. Sens. Env. 24:405-418

Pierce, L.L., Running , S.W. and Riggs, G.A. (1990)
Remote detection of canopy water stress in coniferous forests using the NS001 thematic mapper simulator and the thermal infrared multispectral scanner.
Photogrammetric Engineering and Rem. Sensing 56:579-586

Pinty, B., Verstraete, M.M. and Dickinson, R.E. (1989)
A physical model for predicting bidirectional reflectances over bare soil.
Rem. Sens. Env. 27:273-288

Powell, M.J.D. (1965).
A method for minimizing a sum of squares of non-linear functions without calculating derivatives.
Computer Journ. 8:303-307

Powell, M.J.D. (1964).
An efficient method for finding the minimum of a function of several variables without calculating derivatives.
Computer Journ. 7:155-161

Ramakrishna,R., Nemani & Running, S.W. (1989)
Estimation of regional surface resistance to evapotranspiration from NDVI and thermal-IR AVHRR data.
J. App. Met. 28:276-284

Ranson, K.J., Biehl, L.L. and Bauer. M.E (1985).
Variation in spectral response of soybeans with respect to illumination, view and canopy geometry.
Int. J. Rem. Sens. 6:1827-1842.

Ranson , K.J., Lang, R.H., Sun, G., Chauhan, N.S., Cacciola, R.J. and Kilic, O. (1996)
Mapping of boreal forest biomass from spaceborne synthetic aperture radar.
Pers. Comm.

Ranson, K.J., Lang, R.H., Sun, G., Chauhan, N.S., Cacciola, R.J. and Kilic, O. (1996).
Mapping of boreal forest biomass from spaceborne synthetic aperture radar.
Pers. Comm.

Ranson, K.J., Saatchi, S. and Sun, G (1995).
Boreal forest ecosystem characterisation with SIR-C/XSAR.
IEEE Trans. Geosci. Rem. Sens. 33:867-876

Ranson, K.J.and Sun, G. (1994)
Mapping biomass of a northern forest using multifrequency SAR data.
IEEE Trans. Geosci. Rem. Sens. 32:388-395

Reynolds, J.F. and Acock, B. (1997)
Modularity and genericness in plant and ecosystem models.
Ecological Modelleing 94:7-16

Rich, P.M. (1990)
Characterizing plant canopies with hemispherical photographs.
Rem. Sens. Rev. 5:13-29

Riggs, G.A. and Running, S.W. (1991)
Detection of canopy water stress in conifers using the airborne imaging spectrometer.
Rem. Sens. Env. 35:51-68

Ross, J.K. (1981)
The radiation regime and architecture of plant stands.
Junk, Hauge.

Ross, J.K. and Marshak, A.L. (1988)
Calculation of canopy bidirectional reflectance using the Monte Carlo method.
Rem. Sens. Env. 24:213-225

Roujean, J.L., Leroy, M. and Deschamps, P.Y. (1992)
A bidirectional reflectance model of the Earth's surface for the correction of remotely sensed data.
J. Geophysical. Res. 97:20455-20468

Running, S.W. (1989)
Estimating terrestrial primary productivity by combining remote sensing and ecosystem simulation.
In: *Remote sensing of biosphere functioning.* (R.J. Hobbs & H.A. Mooney, eds)

Running, S.W. (1991)
Computer simulation of regional evapotranspiration by integrating landscape biophysical attributes with satellite data.
In: *Land surface evaporation: measurement and parameterisation* (T.J. Schmugge and J-C Andre, eds)

Running, S.W., Nemani, R.R., Peterson, D.L., Band, L.E., Potts, D.F., Pierce, L.L. and Spanner, M.A. (1989)
Mapping regional forest evapotranspiration and photosynthesis by coupling satellite data with ecosystem simulation.
Ecology 70:1090-1101

Russell, G., Jarvis, P.G. and Monteith, J.L. (1989)
Absorption of radiation by canopies and stand growth.
In: *Plant canopies: their growth form and function.* (eds. G. Russel, B. Marshall and P.G. Jarvis)
Cambridge University Press

Sage, R.F., Sharkey, T.D. and Seeman, J.R. (1989)
Acclimation of Photosynthesis to elevated CO_2 in five C_3 species.
Plant Physiology 89:590-596

Saldarriaga, J.G. and Luxmoore, R.J. (1991)
Solar energy conversion efficiencies during succession of a tropical rain forest in Amazonia.
J. Tropical Ecol. 7:233-242

Sellers, P.J. (1989)
Vegetation canopy spectral reflectance and biophysical processes.
In: Theory and applications of optical remote sensing, Chapter 8 (ed: G. Asrar).
Wiley

Sellers, P.J., Berry, J.A., Collatz, G.J., Field, C.B. and Hall, F.G. (1992)
Canopy reflectance, photosynthesis and transpiration. III A reanalysis using improved leaf models and a new canopy integration scheme.
Rem. Sens. Env. 42:187-216

Sellers, P.J., Hall, F.G., Asrar, G. Strebel, D.E. and Murphy, R.E. (1992)
An overview of the first international satellite land surface climatology project (ISLSCP) field experiment (FIFE).
J. Geophysical Research 97:18345-18371

Sequeira, R.A. Olson. R.L. and McKinion J.M. (1997)
Implementing generic, object-oriented models in biology.
Ecological Modelleing 94:7-16

Sinclair, T.R. (1991)
Canopy carbon assimilation and crop radiation use efficiency dependence on leaf nitrogen content.
In: *Modelling crop photosynthesis – from biochemistry to canopy. (Crop science society of America, No.19).* (eds. K.J. Boote and R.S. Loomins)
Crop society of America.

Sinclair, T.R. and Shiraiwa, T. (1993)
Soybean radiation use efficiency as influenced by non-uniform specific leaf nitrogen distribution and diffuse radiation.
Crop Sci. 33:808-812

Spendley, W., Hext, G.R. and Himsworth, F.R. (1962).
Sequential application of simplex designs in optimization and evolutionary operation.
Technometrics 4:441-541

Spitters, C.J.T., Toussaint, H.A.J.M. and Goudriaan, J. (1986).
Separating the diffuse and direct components of global radiation and its implications for modelling canopy photosynthesis.
Agric. For. Met. 38:217-229

Squire, G.R., Marshall B., Terry, A.C. and Monteith, J.L. (1984)
Response to temperature in a stand of pearl millet. VI Light interception and dry matter production.
J. Exp. Bot. 35:1233-1240

Stokes, G.C. (1862).
On the intensity of the light reflected from or transmitted through a pile of plates.
Proc. R. Soc. Lond. 11:545-556

Strahler, A.H. and Jupp, D.L.B. (1990a)
Geometric-optical modelling of forest scenes composed of three-dimensional discrete objects.
In: *Photon-vegetation interactions* (eds: R.B. Myneni and J. Ross)
Springer-Verlag

Strahler, A.H. and Jupp, D.L.B. (1990b)
Modelling bidirectional reflectance of forest and woodlands using boolean models and geometric optics.
Rem. Sens. Env. 34:153-166

Suits, G.H. (1983)
Extension of a uniform canopy reflectance model to include row effects.
Rem. Sens. Env. 13:113-129

Suits, G.H. (1972a)
The calculation of the directional reflectance of vegetative canopies.
Rem. Sens. Env. 2:117-125

Suits, G.H. (1972b)
Azimuthal variation in the directional reflectance of vegetative canopies.
Rem. Sens. Env. 2:175-182

Sun, G., Simonett, D.A., and Strahler, A.H. (1991)
A radar backscatter model for discontinuous coniferous forests.
IEEE Trans. Geosci. Rem. Sens. 29:639-650

Szu, H. and Hartley, R. (1987)
Fast simulated annealing.
Phys. Letters. A:122:157-162

Tanese, R. (1987)
Parallel genetic algorithms for a hypercube.
In: *Proceedings of the second international conference on genetic algorithms.* pp 177-183
Lawerence Erlbaum

Tanese, R. (1989)
Distributed genetic algorithms.
In: *Proceedings of the third international conference on genetic algorithms.* pp 434-439
Morgan Kaufmann

Tanré, D. Deroo, C., Dahaut, P. (1990)
Description of a computer code to simulate the satellite signal in the solar spectrum: the 5S code.
Int. J. Rem. Sens. 11:659-668
Thornley, J.H.M. and Johnson, I.R. (1990)
Plant and crop modelling.
Clarendon Press.

Tooming, H.G. and Gulyaev, B.E. (1967)
Methods of measuring photosynthetically active radiation. (In Russian)
Nauka, Moscow.

Tucker, C.J. (1977)
Spectral estimation of grass canopy variables.
Rem. Sens. Env. 6:11-26

Tucker, C.J. (1979)
Red and photographic infrared combinations for monitoring vegetation
Rem. Sens. Env. 8:127-150

Tucker, C.J. and Sellers P.J. (1986)
Satellite remote sensing of primary productivity.
Int. J. Rem. Sens. 7:1395-1416

Turner, J.R.G. (1987)
Random genetic drift.
In: *The probabilistic revolution* (Eds. L. Kruger, G. Gigerenzer and M.S. Morgan)
MIT Press

Ulaby, F.T., Moore, R.K., and Fung, A.K. (1982a)
Microwave remote sensing, active and passive, vol II. *Radar remote sensing and surface scattering emission theory.*
Artech House.

Ulaby, F.T., Moore, R.K., and Fung, A.K. (1982b)
Microwave remote sensing active and passive, vol III. *From theory to applications.*
Artech House.

van Zyl, J.J. (1993).
The effects of topography on radar scattering from vegetated areas.
IEEE Trans. Geosci. Rem. Sens. GE31:153-160

Vanderbilt, D. and Louie, S.G. (1984)
A Monte Carlo simulated annealing approach to optimization over continuous variables.
Journ. Comp. Phys. 56:259-271

Varlet-Grancher, C., Gosse, G., Chartier, M., Sinoquet, H. Bonhomme, R., and Allirand, J.M. (1990)
Mise au point: rayonnement soilaire absorbé ou intercepté par un couvert végétal.
Agronomie. 9:419-439 (In French)

Verhoef, W. (1984).
Light scattering by leaf layers with application to canopy reflectance modelling: the SAIL model.
Rem. Sens. Env. 16:125-141

von Caemmerer, S., Evans, J.R., Hudson, G.S. and Andrews, T.J. (1994)
The kinetics of ribulose 1,5 bisphosphate carboxylase/oxygenase in vivo inferred from measurements of photosynthesis in leaves of transgenic tobacco.
Planta 195:88-97

Vose, M.D. and Liepins, G.E. (1991)
Punctuatd equilibria ingenetic search.
Complex Systems 5:31-44

Wang, Y.P. and Jarvis, P.G. (1988)
Description and validation of an array model – MAESTRO.
Agric. For. Met. 51:257-280

Wang, Y.P. and Jarvis, P.G. (1988)
Mean leaf angles for the ellipsoidal inclination angle distribution.
Agric. For. Met. 43:319-321

Wang, Y.P., Jarvis, P.G. and Taylor, C.M.A. (1991)
PAR absorption and its relation to above-ground dry matter production of Sitka spruce.
J. Appl. Ecol. 28:547-560

Wang, Y-P., McMurtire, R.E. and Landsberg, J.J. (1992)
Modelling canopy photosynthetic productivity.
In: *Crop photosynthesis: spatial and temporal determinants* (eds. N.R. Baler and H. Thomas.
Elsever Science

Wanner, W. Li, X. and Strahler, A.H. (1995a)
A new class of geometric-optical semi-empirical kernels for global BRDF and albedo monitoring.
Proc. Int. Geosci. Rem. Sens. Symp. 95:15-17

Wanner, W. Li, X. and Strahler, A.H. (1995b)
On the derivation of kernels for kernel-driven models of bidirectional reflectance.
J. Geophysical Res. 100:21077-21089

Waring, R.H. and Franklin, J.F. (1979)
Evergreen coniferous forests of the Pacific Northwest.
Science 204:1380-1386

Watson, D.J. (1937)
The estimation of leaf area in field crops.
J. Agric. Sci. 27:474-483

Weiss, A. and Norman, J.M. (1985)
Partitioning of solar radiation into direct and diffuse visible and near-infrared components.
Agric. For. Met. 34:205-213

Welles, J.M. (1990)
Some indirect methods of estimating canopy structure.
Rem. Sens. Rev. 5:31-43

Weyers, J.D.B., Lawson , T. and Peng, Z.Y. (1997)
Variation in stomatal characteristics at the whole-leaf level.
In: *Scaling up from cell to landscape*, SEB seminar series 63 (eds: P.R. van Gardingen, G.M. Foody and P.J. Curran).
Cambridge University Press

Whitley, D. (1993)
An executable model of a simple genetic algorithm.
In: *Foundations of genetic algorithms 2*. (Ed. D. Whitley)
Morgan Kaufmann

Whitley, D. Das, R. and Crabb, C. (1992)
Tracking primary hypercube competitors during genetic search.
Annals of Mathematics and Artificial Intelligence 6:367-388

Wigneron, J-P., Calvet, J-C., Kerr, Y.H., Chanzy, A., Lopes, A. (1992)
Microwave emission of vegetation: sensitivity to leaf characteristics.
(*Personal communication*)

Wigneron, J-P., Kerr, Y.H., Chanzy, A. and Jin, Y-Q. (1992)
Inversion of surface parameters from microwave measurements over a soybean field.
(*Personal communication*)

Wright, S (1932)
The roles of mutation, inbreeding, crossbreeding and selection in evolution.
*Proc. 6th Int. Congr. On Genetics*356-366

Yamada, N. and Fujimura, S. (1988).
An mathematical model of reflectance and transmittance of plant leaves as a function os chlorophyll pigment content.
in *Proc. Int. Geosci. and Rem. Sens. Symp.* (IGRASS '88), Edinburgh. 13-16 September 1988. pp 833-834

Appendix A

The SAIL Model

In the SAIL model, the radiation regime within the canopy os described by four differential equations as follows:

$$\frac{dE_+}{dx} = -aE_+ + \sigma E_- + s'E_s \tag{A.1}$$

$$\frac{dE_-}{dx} = -\sigma E_+ + aE_- - sE_s \tag{A.2}$$

$$\frac{dE_s}{dx} = kE_s \tag{A.3}$$

Where: E_s is the downward direct solar flux

E_+ is the upward diffuse flux

E_- is the downward diffuse flux

E_0 is the internal radiance in the direction of observation.

k is the extinction coefficient for E_s

K is the extinction coefficient for internal radiance

a is the attenuation coefficient for diffuse flux

$\sigma, \sigma', s, s', u, v, w$ are various scattering coefficients corresponding to the different combinations of incident and scattered radiation flux, as shown in **Table 3.1**.

General solutions for these equations may be written as:

$$E_+(x) = Ae^{mx} + Be^{-mx} + CE_s(0)e^{kx} \tag{A.4}$$

$$E_- = hAe^{mx} + h^{-1}Be^{-mx} + DE_s(0)e^{kx} \tag{A.5}$$

$$E_s = E_s(0)e^{kx} \tag{A.6}$$

Where:

$$m = \sqrt{a^2 - \sigma^2}$$

$$h = \frac{a+m}{\sigma}$$

$$C = \frac{s\sigma - s'(k-a)}{m^2 - k^2}$$

$$D = \frac{s'\sigma + s(k+a)}{m^2 - k^2}$$

To find the constant A and B, substitute boundary conditions, $E_s(0)$, $E_-(0)$, $E_+(-x_n)$ where x_n is canopy depth (assume $x_n = 1$) and introduce term $E_0(x)$:

$$\frac{dE_0}{dx} = wE_s + vE_- + uE_+ - KE_0 \qquad \text{(A.7)}$$

Let $J = wE_s + vE_- + uE_+$ equal source function describing the generation of internal radiance and let K equal the extinction coefficient of internal radiance in the viewing direction, then:

$$\frac{dE_0}{dx} = J - KE_0$$

The radiance in the viewing direction is given by: $L_0 = \frac{E_0}{\pi}$

Let ρ_{sun}, ρ_{sky} be the reflectance under only direct and diffuse illumination respectively and consider two conditions:

(a) $E_-(0) = 0$ i.e. no diffuse, only direct radiation at top of canopy then let

$$\rho_{sun} = \pi \frac{L_{sun}}{E_{sun}}$$

(b) $E_s(0) = 0$ i.e. no direct, only diffuse radiation then define the reflectance

$$\rho_{sky} = \pi \frac{L_{sky}}{E_{sky}}$$

Then the total reflectance is given by:

$$\rho_{total} = \frac{E_{sky}\rho_{sky} + E_{sun}\rho_{sun}}{E_{total}} \tag{A.8}$$

Define $$f_{sky} = \frac{E_{sky}}{E_{sun} + E_{sky}} = \frac{E_{sky}}{E_{total}}$$

Then

$$\rho_{total} = \rho_{sky} \cdot f_{sky} + \rho_{sun}\left(1 - f_{sky}\right) \tag{A.9}$$

So use boundary conditions to calculate *A* & *B* as follows (for a one layer canopy):

Upper boundary

(a) $E_s(0) = E_{sun}$, $E_-(0) = 0$ direct radiation only

(b) $E_-(0) = E_{sky}$, $E_s(0) = 0$ diffuse radiation only

Lower boundary

$$E_+(-1) = \rho_s\left(E_1(-1) + E_s(-1)\right) \tag{A.10}$$

Where: ρ_s is the soil reflectance

The radiance $\Delta L(x)$ into the upper hemi-sphere from an infinitesimal layer Δx is given by:

$$\Delta L(\theta_o, \theta_s, \varphi) = \frac{1}{\pi}\left[u(\theta_o)E_+ + v(\theta_o)E_- + w(\theta_o, \theta_s, \varphi)E_s(\theta_s)\right]\Delta x \tag{A.11}$$

Some interception of radiation will occur by other layers. The probability of interception into direction θ_o along distance *x* is:

$$p(x) = e^{Kx} \text{ where: } K = f(\theta_o)$$

The contribution ΔL, of the elemental layer $\Delta x(i)$ in canopy layer *i* to the total radiance leaving the top most layer, is found from above by including a term for the probability that the upper layers do not intercept the radiance from layer *I*:

$$\Delta L^0(x(i)) = \sum_{l=1}^{i-1} e^{-K_l x_l} e^{K_i x(i)} \Delta L(x(i)) \tag{A.12}$$

This equation is integrated over all layers of the canopy and a term for the contribution from the observed fraction of soil is added as follows:

$$L^0 = \int_{-x_1}^{0} e^{-Kx} e^{Kx} \Delta L(x) dx \tag{A.13}$$

$$\Delta L(x) = \left[\begin{array}{l} \frac{u}{\pi} A e^{mx} + \frac{u}{\pi} B e^{-mx} + \frac{v}{\pi} h A e^{mx} + \frac{v}{\pi} h^{-1} B e^{-mx} + \\ \frac{u}{\pi} C E_s(0) e^{kx} + \frac{v}{\pi} D E_s(0) e^{kx} + \frac{w}{\pi} E_s(0) e^{kx} \end{array} \right] \Delta x \tag{A.14}$$

$$\Delta L(x) = \left[\frac{A}{\pi}(u + vh) e^{mx} + \frac{B}{\pi}(u + vh^{-1}) + \frac{1}{\pi}(u + v + w) E_s(0) e^{kx} \right] \Delta x \tag{A.15}$$

Let $E_s(0) = E_{sun}$ and x_1 = the height (depth) of the canopy, then:

$$L^0 = \int_{-x_1}^{0} e^{-Kx} e^{Kx} \Delta L(x) dx$$

Rearranging

$$\begin{aligned} L^0 = {} & \frac{A}{\pi}(u + vh) \frac{1 - e^{-(K+m)x_1}}{K + m} + \frac{B}{\pi}(u + vh^{-1}) \frac{1 - e^{-(K-m)x_1}}{K - m} + \\ & \frac{E_{sun}}{\pi}(uC + vD + w) \frac{1 - e^{-(K+k)x_1}}{K + k} \end{aligned} \tag{A.16}$$

The reflectance contribution of observed fraction of soil is given by:

$$\frac{1}{\pi} \rho_s \{E_-(-x_1) + E_s(-x_1)\} e^K$$

$$= \frac{\rho_s}{\pi} \left\{ hA e^{-(+m)x_1} + h^{-1} B e^{-(K-m)x_1} + (D+1) E_{sun} e^{-(K+k)x_1} \right\} \tag{A.17}$$

Where e^K is the attenuation in the direction of view

Therefore the complete reflectance equation is:

$$\pi L^0 = A(u+hv)\frac{1-e^{-(K+m)x_1}}{K+m} + B(u+h^{-1}v)\frac{1-e^{-(K-m)x_1}}{K-m} + E_{sun}(uC+vD+w)\frac{1-e^{-(K+k)x_1}}{K+k} + \rho_s\left[hAe^{-(K+m)x_1} + h^{-1}Be^{-(K-m)x_1} + (D+1)E_{sun}e^{-(K+k)x_1}\right] \quad \text{(A.18)}$$

Where the constants A and B are determined from the boundary conditions for

a) $E_{sky} = 0$

b) $E_{sun} = 0$

where in each instance the result is L^0_{sky} and L^0_{sun} :

$$\rho_{sun} = \frac{\pi L_{sun}}{E_{sun}} \quad \rho_{sky} = \frac{\pi L_{sky}}{E_{sky}} \text{ and } \rho_{total} = \rho_{sky} \cdot f_{sky} + \rho_{sun}(1-f_{sky}) \quad \text{(A.19)}$$

Determination of constants *A* & *B*

Condition 1:

$E_{sun} = 0$, $E_{sky} = 1$

Let canopy height be 1, therefore at top of canopy $x = 0$, and at bottom of canopy $x = -1$

From A.5 $$hA + h^{-1}B - E_{sky} = 0 \quad \text{(A.20)}$$

From A.4 and soil boundary condition A.10

$$Ae^{-m} + Be^{m} = \rho_s hAe^{-m} + \rho_s h^{-1}Be^{m} \quad \text{(A.21)}$$

From A.20

$$A = \frac{E_{sky} - h^{-1}B}{h} \quad \text{(A.22)}$$

Substituting for A in A.21

$$\left(\frac{E_{sky}-h^{-1}B}{h}\right)e^{-m}+Be^{m}=\rho_s h\left(\frac{E_{sky}-h^{-1}B}{h}\right)e^{-m}+\rho_s h^{-1}Be^{m} \tag{A.23}$$

$$\left(E_{sky}-h^{-1}B\right)e^{-m}-\rho_s h\left(E_{sky}-h^{-1}B\right)e^{-m}+hBe^{m}-\rho_s Be^{m}=0 \tag{A.24}$$

$$B=\frac{E_{sky}\left(\rho_s h-1\right)e^{m}}{\left(h-\rho_s\right)e^{m}-\left(h^{-1}-\rho_s\right)e^{-m}} \tag{A,25}$$

From A.20

$$A=\frac{E_{sky}\left(1-\rho_s h^{-1}\right)e^{m}}{\left(h-\rho_s\right)e^{m}-\left(h^{-1}-\rho_s\right)e^{-m}} \tag{A.26}$$

Condition 2
$E_{sky} = 0$, $E_{sun} = 1$

Let canopy height = 1, therefore at top of canopy x = 0, and at bottom of canopy x = -1

From A.5

$$hA+h^{-1}B+DE_{sun}=0 \tag{A.27}$$

From A.5 and A.10

$$\begin{gathered}Ae^{-m}+Be^{m}+CE_{sun}e^{-k}= \\ \rho_s\left[hAe^{-m}+h^{-1}e^{m}B+DE_{sun}e^{-k}+E_{sun}e^{-k}\right]\end{gathered} \tag{A.28}$$

From A.27

$$A=\frac{-DE_{sun}-h^{-1}B}{h} \tag{A.29}$$

From A.28

$$\begin{gathered}\left(\frac{-DE_{sun}-h^{-1}B}{h}\right)e^{-m}+Be^{m}+CE_{sun}e^{-k}= \\ \rho_s\left[h\left(\frac{-DE_{sun}-h^{-1}B}{h}\right)e^{-m}+h^{-1}Be^{m}+DE_{sun}e^{-k}+E_{sun}e^{-k}\right]\end{gathered} \tag{A.30}$$

Rearranging

$$\left(\frac{-DE_{sun}-h^{-1}B}{h}\right)e^{-m}+Be^{-m}-\rho_s h\left(\frac{DE_{sun}-h^{-1}B}{h}\right)e^{-m}-\rho_s h^{-1}Be^{-m}= \quad (A.31)$$
$$\rho_s DE_{sun}e^{-k}+\rho_s DE_{sun}e^{-k}+\rho_s E_{sun}e^{-k}-CE_{sun}e^{-k}$$

$$B\left[(h-\rho_s)e^{m}-(h^{-1}\rho_s)e^{-m}\right]-De^{-m}E_{sun}+\rho_s he^{-m}DE_{sun}= \quad (A.32)$$
$$E_{sun}\left[h(\rho_s(D+1)-C)e^{-k}+De^{-m}(1-\rho_s h)\right]$$

$$B=\frac{D(1-\rho_s h)e^{-m}+h\left[\rho_s(D+1)-c\right]e^{-k}}{(h-\rho_s)e^{-m}-(h^{-1}-\rho_s)e^{-m}}E_{sun} \quad (A.33)$$

From A.29

$$A=\frac{-D(1-h^{-1}\rho_s)e^{m}-h^{-1}\left[\rho_s(D+1)-C\right]e^{-k}}{(h-\rho_s)e^{m}-(h^{-1}-\rho_s)e^{-m}} \quad (A.34)$$

The complete equation for reflectance is obtained by subsitiuting Equations A.25 and A.26 or A.33 and A.34 into A.18 for the two conditions

a) $E_{sun} = 0$, $E_{sky} = 1$ and

b) $E_{sun} = 1$, $E_{sky} = 0$

To give A.18a and A.18b

A.18a and A.18b may then be substituted into A.19 to give the reflectance.

Appendix B

Farquhar Photosynthesis Model

Farquhar and von Caemmerer (1982) describe the rate of CO_2 assimilation as being the minimum of a Rubisco-limited or electron-limit rates:

$$A = \min\left\{\begin{array}{l} V_{c\max}\dfrac{C_i - \Gamma^*}{C_i + K_m} - R_d \\ J\dfrac{C_i - \Gamma^*}{4C_i + 8\Gamma^*} - R_d \end{array}\right\} \quad \text{(B1)}$$

Where: C_i is the internal partial pressure of CO_2 in the chloroplast

Γ^* is the CO_2 compensation partial pressure in the absence of dark respiration

R_d is the rate of dark respiration

V_{cmax} is the maximum rate of carboxylation by Rubisco

J is the electron transport rate

K_m is the Michaelis-Menton coefficient for photosynthesis (oxygenation and carboxylation by Rubisco)

A solution to equation B1 account for the effects of light and temperature on photosynthetic rate, provided that C_i is known. Unfortunately C_i is difficult to measure. However it, photosynthetic rate is related to C_i by:

$$A = g_s\left(C_a - C_i\right) \quad \text{(B2)}$$

$$C_i = C_a - \frac{A}{g_s} \quad \text{(B3)}$$

Where: C_a is the ambient CO_2 concentration of the atmosphere

g_s is the stomatal conductance

Therefore substitution of B3 into B1 replaces C_i with a function of the C_a and g_s.

Rubisco limited rate

From B1:

$$A = V_m\left[\frac{C_a - \frac{A}{g_s} - \Gamma^*}{C_a - \frac{A}{g_s} + K_m}\right] - R_d \tag{B4}$$

Rearranging

$$A\left(C_a - \frac{A}{g_s} + K_m\right) = V_m\left(C_a - \frac{A}{g_s} - \Gamma^*\right) - R_d\left(C_a - \frac{A}{g_s} + K_m\right) \tag{B5}$$

$$AC_a - \frac{A^2}{g_s} + AK_m + A\frac{V_m}{g_s} - A\frac{R_d}{g_s} = V_mC_a - V_m\Gamma^* - R_dC_a - R_dK_m$$

$$A^2 - A\left(g_s\left(C_a + K_m\right) + V_m - R_d\right) + g_s\left(V_m\left(C_a - \Gamma^*\right) - R_d\left(C_a + K_m\right)\right) = 0 \tag{B6}$$

Light limited rate

From B1:

$$A = J\left(\frac{C_a - \frac{A}{g_s} - \Gamma^*}{4\left(C_a - \frac{A}{g_s}\right) + 8\Gamma^*}\right) - R_d \tag{B7}$$

Rearranging

$$A\left(4\left(C_a - \frac{A}{g_s}\right) + 8\Gamma^*\right) = J\left(C_a - \frac{A}{g_s} - \Gamma^*\right) - \left(4\left(C_a - \frac{A}{g_s}\right) + 8\Gamma^*\right)R_d \tag{B8}$$

$$\begin{aligned} &4AC_ag_s + 8A\Gamma^*g_s + JA - 4AR_d - 4A^2 = \\ &g_sJC_a - g_sJ\Gamma^* - 4g_sC_aR_d - 8g_s\Gamma^*R_d \end{aligned} \tag{B9}$$

$$4A^2 - A\left(4C_a g_s - 8\Gamma^* g_s - J + 4Rd\right) = g_s\left(J\Gamma^* + 4C_a R_d + 8\Gamma^* R_d - JC_a\right) \quad \text{(B10)}$$

$$\boxed{A^2 - A\left[g_s\left(C_a - 2\Gamma^*\right) + R_d - \frac{J}{4}\right] - g_s\left[R_d\left(C_a + 2\Gamma^*\right) + \frac{J}{4}\left(\Gamma^* - C_a\right)\right] = 0} \quad \text{(B11)}$$

Therefore, assimilation function requires the following parameters:

g_s - stomatal conductance

R_d - respiration rate

J - electron transport rate

C_a - atmospheric CO_2 concentration

Γ^* - CO_2 compensation point

K_m - Michaelis-Menton coefficient for photosynthesis

V_{cmax} - maximum carboxylation rate

The Michaelis-Menton coefficient for photosynthesis, K_m is given by:

$$K_m = K_c\left(1 + \frac{pO_2}{K_o}\right) \quad \text{(B12)}$$

Where: K_c is the Michaelis-Menton constant for carboxylation by Rubisco

K_o is the Michaelis-Menton constant for oxygenation by Rubisco

The temperature sensitivies for K_c, K_o and V_{cmax} are assumed to follow the Arrhenius-type relationship with no change in activation energy over the temperature range examined. The values of these parameters are therefore expressed relative to their values at 25°C (298.2 K), $K_{c(25)}$, $K_{o(25)}$ and $V_{cmax(25)}$ respectively:

$$K_c = K_{c(25)} \exp\left[\frac{E_c}{(298.2)R}\left\{1 - \frac{298.2}{T_c}\right\}\right] \quad \text{(B13)}$$

$$K_o = K_{o(25)} \exp\left[\frac{E_o}{(298.2)R}\left\{1 - \frac{298.2}{T_c}\right\}\right] \quad \text{(B14)}$$

Where: E_c is the apparent activation energy for K_c.

E_o is the apparent activation energy for K_o.

Therefore K_m is a function of $K_{c(25)}$, $K_{o(25)}$, E_c, E_o, T_c, pO_2

$$V_{c\max} = V_{c\max(25)} \exp\left[\frac{E_v}{(298.2)R}\left(1-\frac{298.2}{T}\right)\right] \tag{B15}$$

Where: E_v is the apparent activation energy for V_{cmax}.

Therefore V_{cmax} is a function of $V_{cmax(25)}$, E_v, T

The electron transport rate J is related to the maximum light-saturated rate of electron transport J_{max} and the absorbed irradiance.

$$J_{\max} = J_{\max(25)} \cdot \frac{\exp\left[\frac{\left(\frac{T}{298.2}-1\right)E_J}{RT}\right]\left\{1+\exp\left(\frac{298.2\cdot S_J - H_J}{298.2R}\right)\right\}}{1+\exp\left(\frac{S_J T - H_J}{RT}\right)} \tag{B16}$$

Therefore J_{max} is a function of S_J, H_J, T, E_J, $J_{max(25)}$

The relationship between J and J_{max} follows a PAR response curve as described by Evans & Farquhar (1991) (see Equation 5.2), which can be rewritten in quadratic form:

$$\Theta J^2 - (I_a\alpha + J_{\max})J + I_a\alpha J_{\max} = 0 \tag{B17}$$

Where: Θ is the convexity of the light response curve

I_a is the absorbed PAR

α is the initial slope of the photosynthetic light response curve.

Therefore, J is a function of Θ, α, I_a, J_{max}

Appendix C

The Omega Coupling Factor

Jarvis and McNaughton coupling factor Ω for leaf evaporation

Amphistomatous leaves

Assume Leaf to be symmetrical, i.e.

$$g_{b1} = g_{b2} = g_b \text{ and } g_{s1} = g_{s2} = g_s$$

Where: g_{b1} is the leaf boundary layer conductance for surface 1

g_{s1} is the leaf stomatal conductance for surface 1

Also, assume $\frac{g_{bH}}{g_{bV}} = 1$ rather than the more probable value of 0.93

Where: g_{bH} is the boundary layer conductance to sensible heat transfer

g_{bV} is the boundary layer conductance to water vapour transport

The equation for latent heat transfer from the (symmetrical, amphistomatous) leaf becomes:

$$\lambda E = \frac{\varepsilon R_n + 2c_p D_a \frac{g_b}{\gamma}}{\varepsilon + 1 + \frac{g_b}{g_s}} \qquad \text{(C1)}$$

Where: D_a is the saturation vapour pressure deficit of the ambient air measured outside the leaf boundary layer

c_p is the molar heat capacity of dry air at constant pressure

ε is the ratio of the slope of relationship between D_a and temperature to the psychrometer constant

γ is the psychrometer constant

R_n is the net radiation flux density

The factor 2 appears in the numerator because the conductances are defined on a single surface area basis.

If leaf boundary layer conductances become very small (leaf is decoupled from the atmosphere), then C1 approaches the limit given by:

$$\lambda E_{equib} = \frac{\varepsilon R_n}{\varepsilon + 1} \quad \text{(amphistomatous)} \tag{C3}$$

If the boundary layer conductance is very large (the ambient saturation deficit is "imposed" at the leaf surface) and C1 tends to:

$$\lambda E_{imp} = \frac{c_p}{\gamma} 2 g_s D_a \tag{C4}$$

Using C3 and C4, C1 can be rewritten as:

$$\boxed{E = \Omega E_{equib} + (1 - \Omega) E_{imp}} \tag{C5}$$

Where Ω is a "coupling" factor is given by:

$$\Omega = \frac{(\varepsilon + 1)}{\varepsilon + 1 + \frac{g_b}{g_s}} \tag{C6}$$

Hypostomatous leaves

For a general hypostomatous leaf with stomata on only one side the equation for latent heat transfer from the leaf can be written:

$$\lambda E = \frac{\varepsilon R_n + c_p D_a \frac{(g_{b1} + g_{b2})}{\gamma}}{\varepsilon + 1 + \frac{(g_{b1} + g_{b2})}{g_{s1}} + \frac{g_{b2}}{g_{b1}}} \tag{C7}$$

Letting g_{b1} and g_{b2} tend to zero as before gives the equilibrium evaporation rate:

$$\lambda E_{equib} = \frac{\varepsilon R_n}{\varepsilon + 1 + \frac{g_{b2}}{g_{b1}}} \quad \text{(hypostomatous)} \tag{C8}$$

Similarly the imposed evaporation rate is:

$$\lambda E_{imp} = \frac{c_p}{\gamma} g_{s1} D_a \tag{C9}$$

The coupling factor is therefore:

$$\Omega = \frac{\left(\varepsilon + 1 + \frac{g_{b2}}{g_{b1}}\right)}{\varepsilon + 1 + \frac{g_{b1} + g_{b2}}{g_s} + \frac{g_{b2}}{g_{b1}}} \tag{C10}$$

Canopy Evaporation

Using the big-leaf approximation the equations for canopies follow directly from the treatment of a single leaf. The equation for latent heat transfer from the canopy can be written:

$$\lambda E_{canopy} = \frac{\varepsilon(R_n + S) + \frac{c_p D_m g_{b(canopy)}}{\gamma}}{\varepsilon + 1 + \frac{g_{b(canopy)}}{g_{s(canopy)}}} \tag{C11}$$

Where the subscript m indicated that the saturation vapour pressure deficit is measured in the mixed layer above the canopy.

The equilibrium and imposed evaporation rates are given by:

$$\lambda E_{equib(canopy)} = \frac{\varepsilon(R_n + S)}{\varepsilon + 1} \tag{C12}$$

$$\lambda E_{imp(canopy)} = \frac{c_p}{\gamma} g_{s(canopy)} D_m$$

Correction using radiation conductance

The above analysis assumes that net radiation R_n does not vary with surface temperature. However, this assumption can lead to significant errors in some cases. The omega formulation can be extended by introducing the concept of isothermal net radiation R_{iso} and a radiative conductance term g_R (see Chapter 5). This leads to (for a symmetrical amphistomatous leaf):

$$\Omega' = \frac{1 + \varepsilon + \frac{g_R}{g_b}}{1 + \varepsilon + \frac{g_b + g_R}{g_s} + \frac{g_R}{g_b}} \tag{C13}$$

The equilibrium evaporation is then given by:

$$\lambda E_{equib} = \frac{\varepsilon R_{iso}}{1 + \varepsilon + \frac{g_R}{g_b}} \tag{C14}$$

The equation for the imposed evaporation (C4) remains unchanged and the Penman-Monteith equation becomes:

$$E = \Omega' E_{equib} + (1 - \Omega') E_{imp} \tag{C15}$$

Similar equations can be derived for hypostomatous leaves or canopies.

Appendix D

Examples of FRAMEWORK Input Files

```
#*********************************************************************
# Environment object initialization file

# air pressure (Pa)
      101300

# Ca, average atmospheric CO2 concentration (Pa)
      36.5   # equivalent to 360 ppm

# Z, height at which wind speed measured
      3.5
```

```
#*********************************************************************
# Evapotranspiration object initialization file

# water_cap_canopy,  water storage capacity of canopy  (mol m-2 leaf area)
      9.4e-12

# water_on_canopt_limit, the threshold value for water_on_canopy above
# which evaporation of intercepted water begins   (mol m-2 leaf area)
      0.0
```

```
#*********************************************************************
# Leafoptics object initialization file
#
# soil optical properties
# -----------------------
# soil reflectance PAR (%)     NIR (%)
             10              30

# PROSPECT model parameters
# -------------------------
# structure chlorophyll       water       protein     cellulose + lignin
# N          cab (µg cm-2)    cw (cm)     cp (g cm-2)    cc (g cm-2)
  1.61        50.8            0.0005       0.001          0.004
```

```
#*********************************************************************
# Respiration object initialization file
#
```

```
# intcpt, intercept on respiration v nitrogen curve
      0.697e-6

# slope, slope of respiration v N curve
      0.0    # no nitrogen response

# calibration temperature for extrapolating respiration v temp relationship
# Tcalib, (Celsius)
      0

# coefficient for extrapolating respiration v temperature relationship
# Tcoeiff
      0.047
```

```
#*****************************************************
# Site object initialization file
# latitude and longitude in degrees n/s and e/w
#
#      Hapex Sahel SSS
#
# latitude    deg    min    sec          N or S
              13     14     48     N

# longitude   deg    min    sec          E or W
              2      214    43     E

# time zone number (1-24)
      1
```

```
#*************************************************************************
# Soilwater object initialization file

# number of soil layers
      3

# water holding capacity of each layer (mols per unit volume)
# and initial water content of each layer (mols per unit volume)

# layer1
# capacity (mols)    initial content
      3400           3400
# layer2
# capacity (mols)    initial content
      3400           3400
# layer3
# capacity (mols)    initial content
      3400           3400

# factor governing probability of free through fall
   0.2  # minimal ftf

# water potential limits (Pa)
#    max      min
     2e6      0.4e6
```

```
#*************************************************************************
```

```
# stom_cond data

# CO2 dependent parameters
# ------------------------
# min and max daily CO2 levels (ppm)
#
#      min           max
       300           360    # min=35.5 Pa; max = 39.5 Pa

# gs_max, maximum stomatal conductance (mol m-2 s-1)
       0.767

# gs_dark, stomatal conductance in the dark (mol m-2 s-1)
       0.178

# temperature dependent parameters
# --------------------------------
#low and high temperature thresholds and optimum temperature (degrees C)
#
# T_low_thres T_hi_thres    T_opt
       -5     40            30

# VPD dependent parameters
# ------------------------
#
# VPD threshold (Pa)
       0.0

# VPD reciprocal coefficient for VPD in Pa  (dimensionless)
       0.094e-3

# light dependent parameters
# --------------------------
# initial slope of light response curve
# units are mols H2O per mol of quanta
#
# value fitted to sitka data - may not be typical
       0.907e-3

# minimum allowed value of gs_water (dimensionless)
       0.25

# soil water dependent parameters
# -------------------------------
# wpot_thres, threshold value for water potential (Pa)
       0.4e6

# wpot_intcpt, value of water potential when gs_water = 0 (Pa)
       2.5e6
```

```
#*****************************************************
# Structure object initialization file
#
#      LAI    canopy height (m)    canopy radius (m)
       0.5    1.5                  1
#
# mean leaf inclination angle, mlang  (degrees)
       67.4   #spherical

# initial biomass
# canopy (kg)        roots (kg)
       1000          500
```

```
# Nconc, nitrogen concentration (g m-2 leaf area)
      14

# partit_coeff, biomass partitioning coefficient
      0.2

# x,y,z co-ordinates of component (optional)
#     x      y      z
      0      0      0

# leaf clumping coefficient (optional)
      0

# height to base of canopy (m)
      0

# specific leaf area (m2 kg-1)
      0

# leaf width (m)
      0.05
```

```
#*****************************************************
# photosynthesis object initialization file (Farquhar Model)
#
# Kc25, M-M constant for carboxylation by rubisco @ 25 C (Pa)
      25.8

# Ko25, M-M constant for oxygenation by rubisco @ 25 C (Pa)
      17.1e3

# GAMMA, CO2 compensation point (Pa)
      3.86


#activation energies in  (J mol-1)
#     Ej           Ec           Eo           Ev
      55.0e3       59.43e3      36.0e3       53.0e3

# alpha
      0.42

# theta
      0.7

# Jmax temperature parameters  (J mol-1)
#     Sj           Hj
      0.49e3       147.0e3

#ambient partial press O2  (Pa)
      21.3e3


# calibration temp for max values of Vc and J  (Celsius)
# ie Vcamx25 and Jmax25 represent the max values of Vc and J at this temp
      25

# Vcmax and Jmax @ calib_temp Celsius - if these params are given values > 0
# they are used.  However, if either Vcmax25 or Jmax25 are 0
# then six further parameters are read in which describe the relationship
# between Vcmax and Nitrogen and Jmax and Nitrogen
```

```
#
#
# Vcmax @ 25 C (mol m-2 s-1)
      0       # read in N relationship (typical value Vcmax25 =  44.0e-6)

# Jmax @ 25 C (mol m-2 s-1)
      0       # read in N relationship (typical value Jmax25  = 134.0e-6)

#########################################################################
# the following are only required if either Vcmax25 or Jmax25 = 0
#
# slope of Jmax v N curve in mol g-1 s-1
      63e-6

# intercept on Jmax v N curve
# ie value of N when Jmax = 0 in g m-2
      0.0

# calibration temp at which Jmax v N relationship formed (Celsius)
      26.5


# slope of Vcmax v N curve in mol g-1 s-1
      21e-6

# intercept on Vcmax v N curve
# ie value of N when Vcmax = 0 in g m-2
      0.0

# calibration temp at when Vcmax v N relationship formed (Celsius)
      26.5
```

Appendix E

Glossary of Models Used

Model:	SAIL
Reference:	Verhof 1984.
Type:	Canopy BRDF model.
Description:	Uses the Kubelka-Munk approximation to the radiative transfer equation. A turbid medium approach is used, therefore the canopy is assumed to horizontally homogeneous. No leaf size or other geometric effects are included so there is no hotspot representation. Predicts the canopy reflectance for a single wavelength given the leaf and soil optical properties for that wavelength.

Model:	PROSPECT - REDUX
Reference:	Jacquemoud and Baret 1990 (PROSPECT).
Type:	Leaf optical properties model.
Description:	Radiative transfer model based on the generalized plate model of Allen *et al* (1969). Uses simple look-up-table approach to model the non-directional spectral reflectance and transmittance of leaves.

Model:	PROSAIL
Reference:	Chapter 4.
Type:	SAIL and PROSPECT linked.
Description:	PROSAIL is a linking of the SAIL and PROSPECT models. All the input parameters of PROSAIL are wavelength independent with the exception of the soil reflectance. If the soil spectral reflectance is known then PROSAIL is able to predict the canopy BRDF for any wavelength from a single set of parameters.

Model:	SOILSPECT
Reference:	Jacquemoud *et al* 1992.
Type:	Soil spectral BRDF model.
Description:	SOILSPECT is based on the work of Hapke (1981). The hotspot effect is included via a soil roughness parameter, h. SOILSPECT requires five other input parameters, four parameters governing the directional reflectance, which are wavelength independent, and the single scattering albedo ω, which is a strong function of wavelength.

Model:	SOILSPECT+
Reference:	Chapter 4
Type:	Soil spectral BRDF model.
Description:	SOILSPECT+ is a modification to SOILSEPCT to include the effects of soil moisture on reflectance. SOILSPECT+ add one extra parameter to the six used by SOILSPECT, that is the soil water layer thickness l. The ω parameter in SOILSPECT and SOILSPECT+ is not wavelength independent. Therefore, in order to these models with multi-wavelength reflectance a look up table is constructed for ω based on a range of soil types and water contents.

Model:	PROSAIL+
Reference:	Chapter 4 & 7
Type:	SAIL and PROSAIL and SOILSPECT+ linked.
Description:	PROSAIL+ is a linking of the SAIL, PROSPECT and SOILSPECT+ models, to form a single model of surface directional spectral reflectance.

Model: ROW

Reference: Goel and Grier (1986a,b)

Type: Canopy BRDF model, incorporating row effects.

Description: ROW is a version of the SAIL model modified to include row planting effects. In ROW the canopy is composed of rows of elliptically shaped canopies. Two parameters are added to describe the canopy shape, J and P where P is the row period and J is the fraction of the distance P covered by half of the plant canopy.

Model: TRIM

Reference: Goel and Grier (1988)

Type: Canopy BRDF model, incorporating row effects.

Description: TRIM is an extension of the ROW model, where the canopy is composed of individual elliptical sub-canopies, occupying a regular grid. Thus, there are effectively two row directions, specified by two axes at right angles to one another. TRIM introduces two further parameters in addition to the ROW model, D and Q, which are equivalent to J and P along the second axis.

Model: Li-Strahler

Reference: Li and Strahler 1985, 1986, 1992

Type: Geometrical-optical canopy BRDF model.

Description: Numerous enhancements have been applied to the original Li-Strahler (1985) nadir–viewing model, including adding of a translucent crown (Li and Strahler 1986) and off nadir viewing (Li and Strahler 1992). The model predicts by directional reflectance using a geometrical-optical approach in which the directional reflectance is modelled as a phenomenon that results when a series of discrete three-dimensional objects (e.g. tree crowns) are illuminated and viewed from different positions.

Model: Bégué

Reference: Bégué 1993

Type: Sparse canopy reflectance model

Description: In the Bégué model the canopy is approximated by an array of porous cylinders representing discrete clumps of vegetation. Mutual shadowing in taken into account, but only single scattering is computed. The model has two sub-models governing the canopy macrostructure and microstructure. The macrostructure submodel adopts a geometric approach to calculate the canopy reflectance based on the cylinder size and shape, whilst the microstructure model uses average transmittance theory applied to each cylinder.

Model: FvC

Reference: Farquhar, von Caemmerer and Berry 1980

Type: Mechanistic, biochemical model of C_3 photosynthesis.

Description: The FvC model describes assimilation as being the minimum of, either the Rubisco-limited, or electron limited assimilation rates. The model has been widely used and there are numerous possible formulations, refer to text for specific details.

Model: CRCB

Reference: Collatz, Ribas-Carbo and Berry 1992

Type: Mechanistic, biochemical model of C_4 photosynthesis.

Description:

The CRCB is an intercellular transport model that links an FvC-type C_3 photosynthesis model with a carbon pump driven by the activity of PEP carboxylase. Carbon, is fixed into C_4 acids in the mesophyll, transported to the bundle sheath cells , and released as CO_2. Leakage of inorganic carbon from the bundle sheath cells back into the intercellular spaces occurs because of the large gradient tin CO_2 concentration created by the metabolic pump.

Model:	FRAMEWORK
Reference:	Chapter 8
Type:	Multi-functional vegetation model.
Description:	The FRAMEWORK model is an object-oriented, highly flexible modular model structure. Individual area of vegetation function such as stomatal control, photosynthesis, bidirectional reflectance and soil water content are represented as individual modules. FRAMEWORK provides the linkage between these modules. Although several default modules are "supplied" within FRAMEWORK, modules are easily inter-changed and replaced.

Model:	P-FRAME
Reference:	Chapter 9
Type:	The FRAMEWORK model modified to run on parallel computer harware.
Description:	P-FRAME is a port of the FRAMEWORK kernel to MPI to enabling running of the program on parallel computers.

Appendix F

FRAMEWORK Default Parameters

Parameter	Value	Units
Photosynthesis		
T_{cal}	25	°C
$K_{c(25)}$	17.1	Pa
$K_{o(25)}$	25.8×10^{6}	Pa
$\Gamma_{(25)}$	3.86	Pa
$V_{cmax(25)}$	44×10^{-6}	$m^{-2}s^{-1}$
$J_{max(25)}$	134×10^{-6}	$m^{-2}s^{-1}$
H_J	147×10^{3}	J mol^{-1}
S_J	0.49×10^{3}	J mol^{-1}
Θ	0.6	
E_c	59.43×10^{3}	J mol^{-1}
E_o	36.0×10^{3}	J mol^{-1}
E_v	53×10^{3}	J mol^{-1}
E_j	41×10^{3}	J mol^{-1}
pO_2	21.3×10^{3}	Pa
α	0.42	-
Stomatal Conductance		
T low threshold	-5	°C
T high threshold	40	°C
T optimum	15	°C
VPD threshold	0.0	Pa
VPD reciprocal coefficient	2.0	Pa
co2 min	390	ppm
co2 max	350	ppm
g_s min	0.5	mol m^{-2} s^{-1}
g_s max	0.01	mol m^{-2} s^{-1}
g_s light min slope	0.0204	
Soil water potential max	2.5	MPa
Soil water potential min	0.4	MPa

Appendix G

MAESTRO Derived Radiation Routines

Direct Radiation Interception

The probability that a light ray will not be intercepted after travelling a distance s through a canopy of randomly distributed foliage elements is given by:

$$\exp(-k_s \mu s) \tag{G.1}$$

Where: μ is the leaf area density (m^2 of one-sided leaf area per m^3 canopy)

k_s is the extinction coefficient, which depends upon the leaf angle distribution (Ross, 1981)

For a canopy of infinite horizontal extent, s is given by:

$$s = \frac{(H - z)}{\cos \eta_r} \tag{G.2}$$

Where: η_r is the angle if incidence if the radiation

H is the height of the canopy

z is the vertical position of the point of interest

Consider a canopy divided into N layers of equal leaf area increment ΔF, the beam interception of the j^{th} layer is given by:

$$I_{B,j} = \exp\left(\frac{-k_s (N - j)\Delta F}{\cos \eta_r}\right) \tag{G.3}$$

Where: F is the cumulative leaf area from the top of the canopy to the point of Interest

So the penetration of beam radiation to any point in the canopy above the j^{th} layer can be expressed as:

$$R_{-,N} f_B I_{B,j} \tag{G.4}$$

Where: $R_{-,N}$ is the total downward radiation above the canopy for a given waveband.

f_B is the beam fraction

Equation G.4 can also be applied to discontinuous canopies such as forests characterised by a number of separate tree crowns. In this case, s must be solved as a function of position within any given tree crown, for a ray incident from any direction, considering the passage of the ray through any adjacent crowns in the canopy. This problem was solved for an array of ellipsoids of any shape by Norman and Welles (1979), but see also extensive treatment by Ross (1981).

Diffuse Radiation Penetration

Diffuse radiation incident of a leaf arises from three sources, (i) the sky, (ii) other foliage, and (iii) the soil. The fraction of diffuse radiation penetrating without interception through a single canopy layer of thickness ΔF can be approximated by:

$$I_D = 2\sum \left\{ \exp \frac{-k_s \Delta F}{\cos \Theta_i} \right\} \sin \Theta_i \cos \Theta_i \tag{G.5}$$

Where: the summation is over all possible incident zenith angles Θ.

The downward diffuse flux above the j^{th} canopy layer originating from both the sky, and other foliage is given by:

$$R_{-,j} = R_{-,J+1}(\tau(1-I_D)+I_D)+R_{+,j}(\rho(i-I_D))+f_B R I_{B,j+1}(1-I_B)\tau \tag{G.6}$$

Where: $R_{-,N}$ is the upward flux above layer j in a given waveband

ρ is the leaf reflectance

τ is the leaf transmittance

The use of the term I_D in equation G.6 implies that ΔF must be small enough so that the probability of leaf overlap within a given layer is small. In practice, this means selecting $\Delta F \approx 0.1$.

$$R_{+,j+1} = R_{+,j}(\tau(1-I_D)+I_D)+R_{-,j+1}(\rho(1-I_D))+f_B R I_{B,j+1}(1-I_B)\rho \tag{G.7}$$

The above equations cannot be applied to any one layer in a single step. Firstly, f_B is set to zero and the ratio $A_{j+1} = \frac{R_{+,j+1}}{R_{-,j+1}}$ is calculated. Then by setting the soil reflectance = A_0, all A_j can be found. Subsequently solving equation G.6 and G.7 yields an expression for all downward diffuse fluxes:

$$R_{-,j} = \frac{R_{-,j+1}[\tau(1-I_D)+I_D]}{[1-\rho(1-I_D)A_j]} \tag{G.8}$$

Substituting G.8 into the right hand side of G.6 and G.7 yields a better estimate of the fluxes expressed on the left-hand side of these equations. These better estimates can then be re-substituted into the right hand side of G.6 and G.7 to yield still better estimates. This iterative process can be continued until equilibrium is reached.

In addition to shortwave fluxes estimates, equations G.6, G.7 and G.8 can be used to obtain estimates of scattered longwave fluxes, however, they do not take into account the emission of radiation by the foliage. Emission is much more important than reflectance at thermal wavelengths, therefore, a negligible error results is longwave reflectance is ignored. Equations G.6 and G.7 are then modified giving:

$$R^l_{-,j} = R^l_{-,j+1} I_D + \sigma T^4_{j+1}(1-I_D) \tag{G.9}$$

$$R^l_{+,j} = R^l_{+,j+1} I_D + \sigma T^4_{j+1}(1-I_D) \tag{G.10}$$

Where the mean leaf temperature for the $(j+1)^{th}$ layer is given by:

$$T^4_{j+1} = \sum_{\theta_l} F_{j,\theta_l} T^4_{j+1,\theta_l} \tag{G.11}$$

Where the summation is over all leaf inclination angles θ_l.

The above equations describe the radiation fluxes averaged over the horizontal plane. This is not the same as the radiation the foliage receives unless all the leaves lie in the horizontal plane. In general, leaves are exposed to a range of radiation fluxes because of their inclination angle distribution. Thus, any property that depends upon incident radiation in a non-linear manner (e.g. photosynthesis) must

take the leaf angle distribution into account when calculating the beam radiation flux for sunlit leaves. The correction term for the effects of the leaf angle distribution is given by:

$$\frac{\cos\theta_l}{\cos\eta_r} \tag{G.12}$$

Careful consideration of photosynthesis requires separate treatment of sunlit and shaded leaves because averaging over a very broad range of flux densities introduces errors because of the non linear nature of the assimilation-light response curve. Standard equations may be used to derive the fraction of sunlit and shaded leaves, see for example the classic work by Ross (1981), and also Myneni and Ross (1991).